High Performance Tree Monte Carlo Applied to Solid and Liquid Water

Dissertation

zur

Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)

vorgelegt der

Mathematisch-naturwissenschaftlichen Fakultät

der

Universität Zürich

von

Mandes Schönherr

aus

Deutschland

Promotionskomitee

Prof. Dr. Jürg Hutter (Vorsitz)

Prof. Dr. Joost VandeVondele (Leitung der Dissertation)

Zürich, 2014

Abstract

The development of large scale multi-processor computers provides the users with a new level of performance. To harvest this potential will require adaptations and developments of algorithms, distributing a huge number of concurrent tasks to distinct processors. The objective of this thesis has been the development and implementation of an optimized Markov chain Monte Carlo (MC) algorithm. Single chain MC simulations are improved in performance to calculate time-independent system properties with long relaxation times of chemical systems in reduced times. The presented Tree Monte Carlo (TMC) algorithm creates new configurations, assuming both possible outcomes of a MC step, and manages them in a binary tree. These configurations and hence multiple Markov chain elements are calculated in parallel, which reduces the simulation time logarithmically with an increasing number of employed processors. Furthermore, the performance is enhanced by accurate estimations of the outcomes of the single MC steps, concentrating the computing power on the a priori unknown Markov chain.

The algorithm is applied to sample water ice systems in the isobaric-isothermal ensemble and to calculate structural and electronic properties using density functional theory (DFT). Temperature dependent dielectric constants are computed with both hybrid and semi-local density functionals. The temperature of the phase transition from the proton disordered hexagonal phase (Ih) to the proton ordered phase (XI) is predicted using the Curie-Weiss law and additionally confirmed by direct simulations close to the transition temperature. Furthermore, isobaric-isothermal calculations of bulk liquid water are performed on the second-order Møller-Plesset perturbation (MP2) level of theory. The obtained density and the radial distribution function are in fair agreement with experimental data, which are also compared to DFT calculations, using semi-local and hybrid functionals, as well as functionals including empirical dispersion corrections.

Zusammenfassung

Die Entwicklung von Hochleistungsrechnern mit Multiprozessor-Architekturen stellt leistungsfähige Werkzeuge bereit um rechenintensive Simulationen durchzuführen. Um dieses Potenzial auszuschöpfen, müssen Algorithmen angepasst bzw. entwickelt werden, die eine Vielzahl von unabhängigen Aufgaben auf die einzelnen Prozessoren verteilen.

Das Ziel dieser Arbeit war die Weiterentwicklung und Implementierung eines Monte Carlo (MC) Algorithmus, welcher einzelne, sehr lange, Markov-Ketten in stark verkürzter Laufzeit produziert um damit zeitunabhängige Systemeigenschaften mit langen Relaxationszeiten für chemische Systeme zu berechnen. Der vorgestellte Tree Monte Carlo Algorithmus erstellt, unter Annahme der beiden möglichen Resultate eines MC Schritts, neue Konfigurationen des simulierten Systems und stellt diese in einem binären Baum bereit. Diese, und damit mehrere Elemente der Markov-Kette, können zeitgleich berechnet werden, was die Laufzeit mit zunehmender Anzahl an Prozessoren logarithmisch verringert. Zudem wird die Leistung durch ein präzises Abschätzen der vorher ungewissen Ausgänge der jeweiligen MC Schritte verbessert. Damit wird die verwendete Rechenleistung auf die resultierende Markov-Kette konzentriert.

Im Rahmen dieser Arbeit, wird dieser Algorithmus für Simulationen von Wasser und Eis unter konstantem Druck und Temperatur, unter Verwendung von Dichte-Funktional-Methoden, angewendet. Die Phasenübergangstemperatur von Eis Ih zu Eis XI wurde über das Curie-Weiss Gesetz abgeschätzt und auch durch direkte Simulation eingegrenzt. Berechnungen von Dichten und radialen Verteilungsfunktionen für flüssig Wasser unter Verwendung von Møller-Plesset-Störungstheorie zweiter Ordnung zeigen gute Übereinstimmung mit experimentellen Werten.

Acknowledgement

I would like to thank Professor Jürg Hutter for giving me the opportunity to do my PhD in his group and participate to the development of CP2K, which is in my opinion a great program.

It difficult to express my gratitude for my supervisor Professor Joost VandeVondele. His support, encouragement and motivation guided me across the years of my whole PhD, from the beginning to the end. Without his advice and profound knowledge in molecular simulation and physics of materials, this work would not have been possible. His enthusiasm is infectious and he always provided me an another point of view. I very much enjoyed to collaborate with him.

I would like to thank my colleges Yannick and Mauro for all support and inspiring discussions, their knowledge in quantum chemistry and computer science were a great help. I will miss the pleasant atmosphere in our office and all the jokes.

Furthermore, I thank all the present and former PostDocs and PhD students of my group whose shared the ups and downs of life as PhD student.

Contents

Chapter 1

Introduction

Computer simulations are increasingly used to study chemical and physical systems and complement experiments. Results can give insights in the structure of materials that are difficult or impossible to probe or measure experimentally. The accuracy mainly depends on the underlying theory, raising the computational expenses dramatically with increasing system sizes and the complexity of models. Accurate chemical simulations often employ ab initio methods, such as Density Functional Theory (DFT)[1], utilizing an increased order of basis sets, which increases the computational costs prohibitively[2]. Thus, on one hand, the needs arise for enhanced accuracy and enlarged system sizes, introducing a raising computational effort which has to be performed in reasonable simulation times. On the other hand, available computing power increases rapidly, including already existing petascale computing systems, and exascale systems planned for the near future (approximately 2018, see www.top500.org). The high performance of most recent computational machinery is mainly reached by an increasing number of processors, instead of an increasing speed of the single processor. This change in computer architecture requires adaptations and developments of algorithms, dealing with parallelization techniques, which distribute separate simulation tasks to distinct processors.

Common parallelization strategies distribute the simulated systems, respectively data, to available processors (called domain decomposition). Thus

the multiple parts of the calculation are performed in parallel, reducing the simulation time. However, the distributed parts of the system become smaller and smaller, reaching a lower limit. Furthermore, information has to be communicated between the processors (exchanged between the distributed system parts), thus the total communication increases with the number of processors. As a result, the performance improves with the employed number of processors up to a certain point, afterwards the communication dominates the calculation, slowing down the performance. There exist an optimal number of processes for which time is minimal as shown in Fig. 1.1.

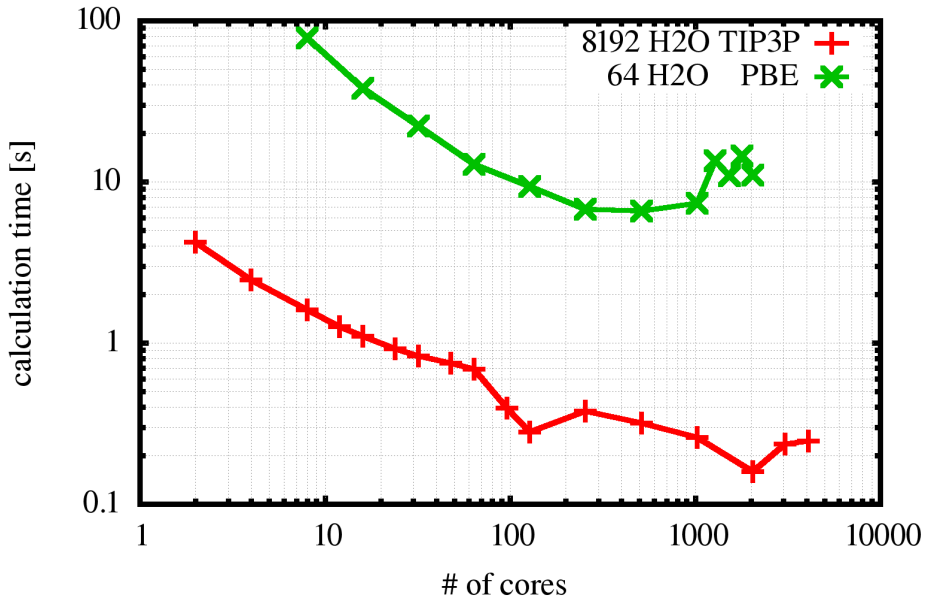


Figure 1.1: The time for a single energy calculation for various parallelization sizes for 8192 H_2O using flexible TIP3P potential (red), 64 H_2O using DFT (PBE functional) (blue). The cost for the methods also varies with the selected parameters, which are not specified in detail here.

The Monte Carlo (MC) method introduces the possibility to compute multiple states of the system in parallel. Simulations of chemical systems mostly employ the Metropolis MC method[3], which modify in each step the previous configuration. As a result, a random walk in configurational space is created, which is called Markov chain. Principally, these MC steps

are intrinsically serial, but since MC is a statistical method, steps can be completely decoupled. This huge potential of parallelization is exploited in a wide variety of fields, e.g. bioinformatics[4], astrophysics[5] or financial economics[6]. The requirements can differ, for example, in the number of particles, the computational costs, the types of moves applied and the coupling between configurations, depending on the simulated systems and applied models. Common parallel MC methods create multiple Markov chains simultaneously, with decreased length at constant computational effort, but all chains generally include calculations of an equilibration phase. Furthermore, system properties with long relaxation times require calculations of long Markov chains. Hence an algorithm parallelizing a single Markov chain is more efficient (compress better the simulation time) compared to the mentioned parallel MC method. Different methods are discussed in more detail later, in Sec. 2.2.

In the present study, the performance of very accurate simulations of molecular systems with long relaxation times is improved to obtain time-independent properties. The investigated high-dimensional problems require expensive energy calculations, utilizing accurate functionals, which significantly dominate the calculation time. The creation of new Markov chain elements are intrinsically coupled to the last element, because complete randomly created configurations result most likely in unphysical states, considering simulations of chemical systems. These configurations are most likely rejected. However, new elements can be created without any force or energy calculation, which is the key point for the later described algorithm. Furthermore, the long relaxation time of the system properties necessitate the sampling of one (very) long Markov chain. Hence, simultaneous calculations of multiple consecutive Markov chain elements (including computational expensive energies) are necessary to reduce the calculation time.

The presented Tree Monte Carlo (TMC) algorithm introduces a parallelization of Markov chain Monte Carlo methods, utilizing massive parallel computers. Multiple elements of the resulting Markov chain are computed in parallel, which significantly reduces the overall simulation time. The algorithm creates possible Markov chain elements, which are handled in a binary

tree, assuming both possible outcomes (acceptance and rejection) of a single MC steps. These elements are computed in parallel on distinct computing units. The efficiency of the presented algorithm is improved by accurate estimations of the a priori unknown Markov chain directions, as well as configurational pre-sampling, and specialized moves.

After this introduction, the basic MC framework will be summarized and common parallelization techniques compared. Then, a detailed description of the general TMC algorithm is presented, using the binary tree representation. The third chapter will treat the implementation of the algorithm, including the optimizations, matching the requirements of molecular simulations. In the second part of this work, the implementation will be verified. The algorithm is applied on accurate simulations of liquid and solid water systems, calculating structural and electronic properties, which are major challenges in condensed matter simulations. Finally all results will be recapitulated and an outlook presented.

Part I

Theory and Implementation

Chapter 2

Theory

This chapter reviews the basic theoretical concept of Markov chain Monte Carlo (MC) methods. The theory is presented for simulations of ensembles with a constant number of particles and constant temperature (NVT and NPT). However the following Tree Monte Carlo (TMC) algorithm is well suited to handle also other ensembles. Here, in particular, single MC steps will be discussed, since the handling of them plays a key role in the TMC algorithm, which creates new configurations considering both possible outcomes of these steps. Advantages and disadvantages of existing parallelization approaches are then discussed. Finally, the TMC algorithm is described in detail, using the binary tree representation. The basic concepts of the implemented optimizations are explained, including pre-sampling, parallel energy calculations, predictions of the Markov chain directions, and specialized moves.

2.1 Monte Carlo

Metropolis Monte Carlo [7] methods are derived from statistical mechanics, producing a random walk in configurational space (Markov chain). Time-independent system properties are calculated by averaging the properties of the resulting Markov chain elements. The sampled points in space (called configurations) are defined by the coordinates of all N particles, $\mathbf{r}$, and, in

general descriptions, their corresponding momenta $\mathbf{p}$. Considering a three dimensional space, $\mathbf{r}$ and $\mathbf{p}$ are elements in $3 \times N$ dimensional space. In isolated systems, the total energies of the constituent particles are defined by the sum of kinetic and potential energy $\mathcal{H} = \mathcal{K} + \mathcal{U}$. This function is called Hamiltonian $\mathcal{H}(\mathbf{r}, \mathbf{p})$.

Applying the MC method, the configurations are Boltzmann weighted with respect to their energies. In other words, the system is found around a point in space $(\mathbf{r}, \mathbf{p})$ with the probability density

$$\mathcal{N}(\mathbf{r}, \mathbf{p}) = \frac{\exp[-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})]}{\int d\mathbf{p}\ d\mathbf{r}\ \exp[-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})]}, \tag{2.1}$$

considering a canonical ensemble. The denominator (usually named Z) derives from the partition function

$$Q = c \int d\mathbf{p}\ d\mathbf{r}\ \exp[-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})], \tag{2.2}$$

where $\beta = 1/k_B T$ and c is a constant of proportionality, see also [7]. Thus, the system properties are averaged over the sampled points with respect to their weight, as they are derived in statistical mechanics (see also [8]),

$$\langle \mathcal{A} \rangle = \frac{\int d\mathbf{p}\ d\mathbf{r}\ \exp[-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})\ \mathcal{A}(\mathbf{r}, \mathbf{p})]}{\int d\mathbf{p}\ d\mathbf{r}\ \exp[-\beta \mathcal{H}(\mathbf{r}, \mathbf{p})]}, \tag{2.3}$$

where the considered properties of the actual sampled configurations $\mathcal{A}(\mathbf{r}, \mathbf{p})$ are weighted with the probability of find the system in the considered states.

This continuous formulation can be approximated for the discrete sampling of a computer simulation as

$$\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^{L} n_i A(\mathbf{r}_i, \mathbf{p}_i), \tag{2.4}$$

reaching the correct distribution of Eq. 2.3 with the law of large numbers in the total number of sampled configurations L. The employed weight $n_i = \mathcal{N} L$ is related to the average number of generated points in a unit volume

around $(\mathbf{r}, \mathbf{p})$. It should be noted that the exact weights are unknown, but due to the specific sampling described by Metropolis and Rosenbluth [7], configurations are sampled with respect to their probability of appearance. As a result a certain state of the system occurs multiple times in the sampling. In contrast to Metropolis and Rosenbluth, here the multiplicity of a single configuration i is taken into account using the weighting factor n_i. The Markov chain Monte Carlo simulations create random walks in space, where each configuration is counted several times, proportional to the probability of the state (Eq. 2.1). As a result, a new configuration $(\mathbf{r}', \mathbf{p}')$ is accepted with a probability equal to the ratio of the probability density of the new and the old configuration $\frac{\mathcal{N}(\mathbf{r}',\mathbf{p}')}{\mathcal{N}(\mathbf{r},\mathbf{p})}$. The resulting Markov chain elements contribute to the investigated properties utilizing Eq. 2.4. In further notations, the points in configurational space are denoted as $\theta = (\mathbf{r}, \mathbf{p})$.

2.1.1 Monte Carlo step

A Monte Carlo step connects two neighboring elements of the Markov chain. Starting with a previously accepted configuration θ_o with the related energy $\mathcal{H}(\theta_o)$, a new configuration θ_n is obtained performing the following three steps:

1. configurational change $\theta_o \rightarrow \theta_n$
2. calculation of $\mathcal{H}(\theta_n)$
3. acceptance check, $r < \mathrm{acc}(\theta_o \rightarrow \theta_n)$

The acceptance criterion is derived from the ratio of the probability density (eq.2.1) of the new and the last accepted configuration. For a canonical ensemble the acceptance criterion reduces to

$$\mathrm{acc}\,(\theta_o \rightarrow \theta_n) = \exp(-\beta[\mathcal{H}(\theta_n) - \mathcal{H}(\theta_o)]) < 1 : w. \tag{2.5}$$

Hence, the new configuration θ_n should be accepted with the probability $\mathrm{acc}\,(\theta_o \rightarrow \theta_n)$. Indeed, this probability is equal to the probability of generating an uniform distributed random number r (obtained from the interval

$[0, 1]$) which is smaller than $\exp(-\beta[\mathcal{H}(\theta_n) - \mathcal{H}(\theta_o)])$. In other words, θ_n is accepted if $r < \mathrm{acc}\,(\theta_o \to \theta_n)$ and rejected otherwise. In case of rejection, the last accepted configuration is counted again, and the next configurational change is based on the last accepted configuration. Hence, a sequence of distinct configurations is created, where each configuration is weighted by the number of following rejected configurations. This weight describes the probability density of this configuration, taking the total number of configurations into account. It should be noted that the acceptance criteria vary for specialized moves, which is shown in the related section 3.4.

The correct sampling is imposed by applying the detailed balance condition, which requires the number of transitions from θ_o to θ_n equal to the number of transitions from θ_n to θ_o, thus

$$\mathcal{N}(\theta_o)\pi\,(\theta_o \to \theta_n) = \mathcal{N}(\theta_n)\pi\,(\theta_n \to \theta_o) \tag{2.6}$$

where π describes the transition probability $\pi = \alpha(\theta_o \to \theta_n) \times \mathrm{acc}(\theta_o \to \theta_n)$. Originally Metropolis chose α to be symmetric ($\alpha(\theta_o \to \theta_n) = \alpha(\theta_n \to \theta_o)$). Manousiouthakis *et al.* showed, that the detailed balance condition can be reduced to the necessary weaker balance condition, preserving the desired distribution [9], which is not treated in detail here. Furthermore the sampling should be ergodic, therefore it has to be ensured that every configuration can be reached in a finite number of steps from every other configuration [8]. A huge variety of configurational changes exists. Depending on the sampled system and the desired system properties, specialized moves speed up the convergence, and spread the sampled configurational space. In Sec. 3.4 a selection of moves and its implementation are described in more detail.

2.2 Parallelization approaches

Nowadays, computer architectures are produced employing an increasing number of processors, already reaching millions of cores (processors). The reason is the ultimately bounded speed of a single processor, reaching the limitations of miniaturization at an atomistic level. As a result, algorithms are

adapted and developed to exploit the increasing computing power available at parallel machine architectures, distributing separate computing tasks to this increasing number of processors. Necessary communication between the distributed tasks are accomplished utilizing the Message Passing Interface (MPI) [10].

The statistical Monte Carlo methods are well suited for parallelization, because configurations can be decoupled and calculated simultaneously. Hence, the simulation time can be reduced utilizing supercomputers of increasing sizes. The most popular parallel Monte Carlo method [11, 12] duplicates the initial configuration on different computing units, creating multiple Markov chains. Hence, compared to a serial algorithm, the number of created Markov chain elements increases linearly with the number of utilized processors. However, the configurations sampled are localized in a configurational space around the initial configuration(s). This localization depends on the size of the configurational changes and the number of MC steps. Multiple distinct initial configurations could spread the sampled configurational space, but in case of a complex chemical system, the creation of distinct configurations near the equilibrium is difficult. Assuming a non-equilibrated starting point, the simulation has to equilibrate first, resulting in a significant change of the obtained system properties. Depending on the system and the investigated properties, fluctuations around the average value remain after the equilibration phase. Thus, an early point for a finalized equilibration phase is difficult to determine during the run time, especially if fluctuations are large in amplitude. Additionally, since each Markov chain has to equilibrate first, the computational cost for calculations of the equilibration phase scales with the number of parallel created chains. Furthermore, the Markov chains need to be sufficiently long to overcome possible energy barriers and reaching all states of high probability. Calculations of slowly converging system properties, e.g. the dielectric constant, require the sampling of an enlarged configurational space. This means that simulations of one long trajectory give more accurate results at the same computational cost than simulations with many short trajectories and are thus preferred to the latter.

Another algorithm, namely Parallel tempering, creates multiple Markov

chains at various temperatures and distributes them to the available computing units[13]. Sampling at high temperatures is used to promote transitions among different regions of high probability density, while low-temperature sampling increases the precision in local regions. Both are combined by configurational swapping, described in more detail in Sec. 3.4.3. Indeed, the computational costs increases for simulations of single temperature properties, employing one of the multiple calculated Markov chains. However, the sampling of states of high probability, which are separated by large energy barriers, can be improved by this technique.

Speculative Monte Carlo methods decrease the run-time by assuming the outcomes of single MC moves and calculating possible configurations simultaneously. Bryd [14, 15, 16] implemented an algorithm on Shared Memory Processing (SMP) architectures, pre-calculating sets of configurations for rejection cases. Hence, in the best case, one configuration is found in each cycle, assuming that enough cores are available. Furthermore, configurations for both possible outcomes (accepted and rejected) of a MC step can be predicted, permitting an increased efficiency, compared to the algorithm described by Byrd. Brockwell [17] presented the "pre-fetching" algorithm, which is described in detail in a theoretical approach. This algorithm is comparable to the basic concepts of the later described TMC algorithm. Theoretically, the algorithm scales logarithmically with the number of utilized processors. The TMC algorithm improves the performance additionally by accurate acceptance probability estimations, through heuristics, and record keeping. The available computing power is concentrated on the a priori unknown Markov chain, reducing significantly the simulation time, discussed in more detail later in Sec. 2.3.5. Mostly, the single tasks (e.g. energy calculation) can be additionally parallelized, employing domain decomposition methods, which will be not considered in detail.

2.3 Tree Monte Carlo

Tree Monte Carlo (TMC) is a speculative approach, specialized for high performance computing of Markov chains. It aims to decrease the time to

solution, employing the increasing computing power of parallel nature, which is becoming more and more dominant. Multiple elements of one long Markov chain are computed simultaneously, assuming the computational expenses for energy calculations to be significantly higher compared to the expenses for configurational change. A master-worker scheme is implemented, dividing the available computing power into worker groups, which additionally distribute the tasks within the groups, utilizing domain decomposition methods.

In this section, first the configurational handling in a binary tree notation will be described in detail. Furthermore, an additional global tree (also a binary tree) necessary for the implemented version of parallel tempering is introduced. Utilizing parallel tempering, N (sub-)trees are created for N temperatures, which are connected by configurational swaps. The global tree elements handle references to the configurations in the other trees, keeping record of configurational swaps. Then, the pre-sampling will be described, speeding up the calculation by preparing configurations, especially for computational expensive configurational changes. Later, simultaneous energy calculations of multiple Markov chain elements will be discussed. The parallel efficiency is improved by concentrating the available computing power on the a priori unknown Markov chain elements, utilizing predictions of the Markov chain directions, which will be described in Sec. 2.3.5. Furthermore, specialized moves improve the convergence of the desired system properties by increasing configurational changes, described in Sec. 2.3.6.

2.3.1 Tree notation

Each Monte Carlo step results in one of two possible outcomes, acceptance or rejection of the created configuration. The Tree Monte Carlo algorithm assumes both possible cases and creates configurations for a variety of possible Markov chains. Since the configurational change can be performed without any energy of force evaluation, all possible configurations of all possible Markov chains can be created. These configurations are represented in a binary tree structure, mapping the assumption of the acceptance check outcomes, shown in Fig. 2.1. Each node represents a possible configuration θ_i

and is connected to two children, assuming either acceptance (green branch) or rejection (red branch) of the actual configuration. The initial configuration θ_0 is, per definition, always accepted and has only an accepted branch.

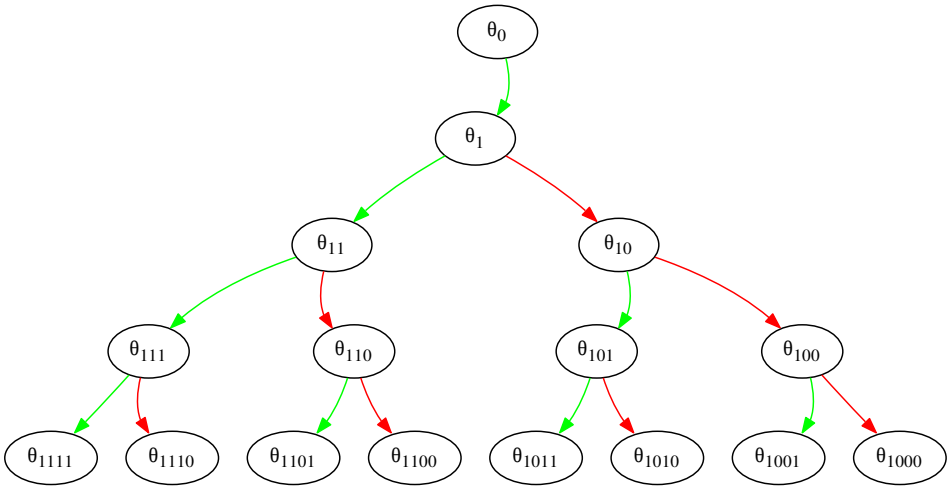


Figure 2.1: Binary tree notation of possible Markov chain elements, written in binary numbering, assuming both possible MC acceptance outcomes. Green branch assumes acceptance, red branch assume rejection

In the following notation, a binary numbering of tree elements is utilized for a better understanding. Thus, the two children of an element θ_i are denoted as: θ_{2i+1} (odd, i binary shifted left plus one) for the accepted branch, and θ_{2i} (even, i binary shifted left) for the rejected branch. Assuming acceptance of the configuration θ_i (e.g. the initial configuration θ_0), the configuration θ_{2i+1} (θ_1) is a configurational change of θ_i (θ_0). Furthermore, assuming rejection of this element θ_{2i+1} (θ_1), the element $\theta_{2(2i+1)} = \theta_{4i+2}$ (θ_{10}) is created, which is another configurational change of θ_i (θ_0). Additional rejection of this element imposes another configurational change of θ_i (θ_0), creating θ_{8i+4} (θ_{100}), and so on, modifying always the last accepted configuration.

In contrast to most other MC implementations, the change is not reverted on the actual element, in case of rejection, since both cases are treated simultaneously. Hence, the configurational basis has to be searched in the tree, related to the applied assumptions. The last assumed accepted configuration θ_a is the parent element θ_i, in case of an odd numbered element θ_{2i+1}, since the parent element is assumed to be accepted. However, in case of an even numbered element θ_{2i}, (multiple) parental elements are assumed to be rejected. Hence the configurational basis has to be searched by following the tree upwards until the next odd numbered element (child on accepted branch) is reached, which assumes acceptance of the parent element θ_a, see Fig. 2.2. This parent element θ_a is the configurational basis and also used

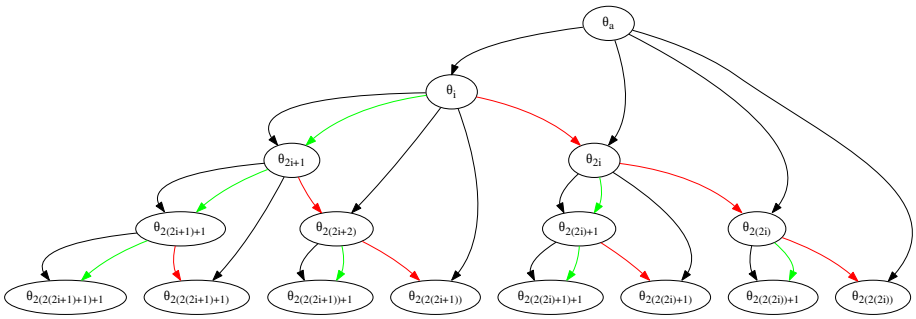


Figure 2.2: Binary tree notation of possible Markov chain elements, written in binary numbering, assuming both possible MC acceptance outcomes. Black arrows originates from the basis elements for configurational change. Hence the two elements connected by the black arrow are compared in the related MC acceptance check.

for the related acceptance check $\mathrm{acc}(\theta_a \to \theta_i)$. The index a can be calculated using

$$a = \begin{cases} \frac{i-1}{2}, & \text{if } i \text{ is odd} \\ \frac{\max\left(j=\frac{i}{2^k}|j,k\in\mathbb{N}\right)-1}{2}, & \text{if } i \text{ is even} \end{cases} \tag{2.7}$$

With the described method a tree of all possible configurations is created. The energies of these configurations can be calculated simultaneously. Af-

terwards, the acceptance check is performed, selecting the "correct" Markov chain elements. Hence, the resulting Markov chain is one way through the tree (also called trajectory), consisting of accepted configurations, weighted by the number of following rejected elements (until reaching the next accepted Markov chain element).

The elements outside this trajectory are discarded, but essential for the parallel implementation. Theoretically, a full binary tree of depth h consists of

$$N = \sum_{i=0}^{h} 2^i = 2^{h+1} - 1 \tag{2.8}$$

elements, but the related Markov chain consists "only" of h elements. Thus, from Eq. 2.8, the theoretical speedup of the algorithm, calculating simultaneously h "correct" Markov chain elements on N computing units, results in an logarithmic behavior

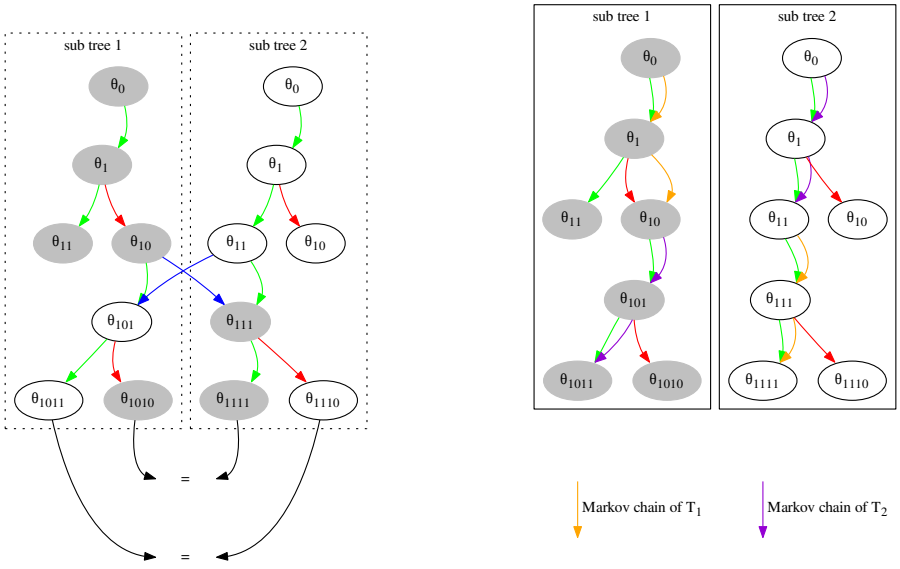
$$h \propto \log_2(N+1). \tag{2.9}$$

The efficiency can be significantly increased by an accurate prediction of the "correct" Markov chain direction. Thus, only the most likely elements are created and the available computing power concentrated on the "correct" Markov chain elements, reducing the elements beside this chain. This optimization is described in more detail in Sec. 3.3.

2.3.2 Global tree notation for Parallel Tempering

Employing the TMC algorithm, configurations are created around a random walk in configurational space (Markov chain). Depending on the simulated system, large energy barriers separate regions of high probability. The Parallel Tempering (PT) method provides the possibility to overcome these energy barriers within a reasonable length of the Markov chain. Multiple Markov chains are sampled on different temperatures including configurational swapping. High temperatures are used to promote transitions between regions in configurational space of high probability density, while the configurational swapping introduces the expanded configurational space into the low temperature sampling. As a result, the efficiency of the sampling on high tem-

peratures is improved by an increased sampling of states of high probability.



(a) swap moves handled in (sub-)trees (b) swap moves handled externally

Figure 2.3: Two different approaches to handle the swap moves are sketched. In the left panel, acceptance checks are performed with a constant temperature within the trees. Thus a swap exchange configurations between the trees. As a result, the swapped configurations have to be copied (the two configurations connected by the blue arrow are equal) equal configurations would be created in both trees ($\theta_{1010} = \theta_{1111}$ and $\theta_{1011} = \theta_{1110}$ and following configurations), assuming the both possible outcomes of the swap move. On the other side (right panel) trees are created independently of temperatures, swaps are handled externally as well as the temperature for the related acceptance checks. As a result, each configuration is created only once (unique) and the resulting Markov chains swap from one tree to another. Gray and white elements represents configurations of two distinct regions of the configurational space.

Assuming all Markov chains start from the same initial configuration (and similar employed random numbers), the Markov chains can be represented by multiple ways through the tree, sampling distinct random walks in configurational space. Thus, N (sub-)trees are created to handle N temperatures.

Without any swap, the "correct" Markov chain of a certain temperature is represented as the resulting trajectory through the related (sub-)tree. But, including swap moves, the Markov chains jump from one region in the configurational space to another. Theoretically, this can be handled in two different ways. On one hand, the configurations could be exchanged between the (sub-)trees (Fig. 2.3a). As a result, further configurations occur in both (sub-)trees, assuming both possible outcomes (acceptance/rejection) of the swap move. This duplication of configurations is avoided by handling the swap move outside the (sub-)tree environment, as well as the utilized temperatures. On the other hand, the possible random walks in configurational space are continued as usual in the (sub-)trees (no configurational exchange, see Fig. 2.3b) and the assignment of elements to the Markov chains (utilized temperatures) is performed in a (so called) global tree. A global tree element (Θ_i and squared nodes in Fig. 2.4) maps a set of configurations (one of each temperature). Pointers to the (sub-)tree elements are stored in a list, where the positions in the list determine the temperature of the related configuration. Hence, a swap move is performed by exchanging two pointers in the list.

Within each step in the global tree, either one configuration of one temperature is modified, or two configurational pointers are exchanged (swap move). The modified configuration is obtained by following the (sub-)tree branch of the previous employed (sub-)tree element with respect to the assumption (acceptance/rejection) of the global tree element, which last modified this (sub-)tree element. For example, the global tree element Θ_{110} (see Fig. 2.4) should be created. The configurational change should be applied on the element of the second temperature. The (sub-)tree element θ_1 (not necessarily on (sub-)tree 2) is last modified at Θ_1. Since Θ_{110} is connected to the acceptance branch of Θ_1, the configuration θ_1 is assumed to be accepted. Hence, the child in accepted direction of θ_1 (θ_{11}) is employed for the desired move. If this desired (sub-)tree element does not already exist, it has to be created first. Within the speculative approach, one (sub-)tree element could be employed in multiple global tree elements and at distinct temperatures. Consequently, possible moves have to be sampled independently of

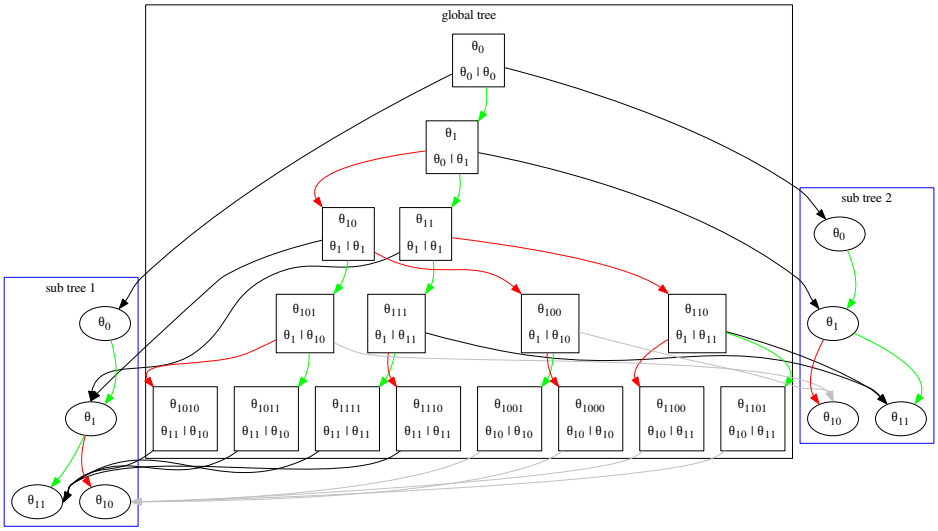


Figure 2.4: Relationship between the elements of a global tree and two (sub-) trees. Nodes with rectangular shape are the global tree elements, while round nodes are (sub-)tree elements, containing the configurational data. All nodes are defined by a binary number (first line in the node). The global tree nodes contain in the second line a list with the related sub tree element numbers (of temperature 1 (right) and temperature 2 (left)). Green and red arrows assume acceptance or rejection, while the black/grey arrows points to the related sub tree element of the actual configurational change.

temperatures.

The global tree trajectory (chain of finally accepted and rejected elements) defines the Markov chain elements of all temperatures. The modified (sub-)tree elements are sorted to the Markov chains of the related temperatures, employing the configurational weighting as described above.

2.3.3 Pre-sampling

The efficiency of Monte Carlo algorithms mainly depends on the configurational change. An increased step size of random particle displacements

results in a decreased acceptance. Therefore, specialized moves are applied to maintain sufficiently large acceptance rates at reasonable large configurational changes, which is discussed in more detail in Sec. 2.3.6. In general, the configurational changes can be performed without any energy or force evaluation. However, the computational costs of these moves can vary significantly, for example when using the Nested Monte Carlo algorithm, which connects two configurations by a Markov chain, based on an approximate potential (described in Sec. 3.4.1). These increased computational costs would result in extended simulation times, performing the move and energy calculation one after each other.

Depending on the complexity, the moves can be parallelized similar to the energy calculations, but usually the parallelization of the energy calculations scales significantly better with the number of employed processors. Therefore, computing units of different sizes should be employed for calculating the configurational change and the energy, respectively. Furthermore, applying the speculative approach of TMC, configurational changes of possible new elements can be performed (simultaneously) on distinct computing units, while calculating the previous configurational energy. When finishing the energy calculation, the next one can start immediately without any delay, reducing the overall calculation time. Theoretically, the pre-sampling can speed up the calculation by a factor of two, assuming the time for configurational change almost equals to the time for energy calculation.

2.3.4 Parallel energy calculations

Single chain Monte Carlo methods are intrinsically serial, creating new configurations by modifying the last accepted ones. In principal, the overall simulation time consists of the time computing the whole number of Markov chain elements, including configurational changes, energy calculations and acceptance checks. The time for performing the acceptance check will be neglected in further considerations, since the acceptance criteria consists only of a comparison of two energies (and a comparison with a random number). In addition, computational expensive moves can be prepared utilizing

the pre-sampling method, described above (Sec. 2.3.3), reducing the cost for configurational changes to a negligible value. As a result, the energy calculations remain the dominant part of the simulation.

The TMC algorithm introduces a method for simultaneous calculation of multiple elements of the a priori unknown Markov chain, reducing the simulation time. Possible Markov chain elements are created in the binary tree structure and their energies are calculated parallel, on distinct computing units. Theoretically, assuming an infinite number of computing units (and already prepared configurations), all energies can be computed simultaneously. Hence, the whole Markov chain could be calculated within the time of one energy calculation, neglecting communication overhead.

For each available computing unit, a tree element at the most probable position in the tree (estimating the "correct" Markov chain direction) is created and computed, hence parts of the tree are computed in parallel. These dynamically created parts consist of Markov chain elements and elements beside the chain, necessary for the parallel implementation. The number of simultaneously calculated Markov chain elements (and hence the speedup) depends on the number of working groups and the accuracy of predicting the Markov chain directions.

2.3.5 Acceptance probabilities (MC and tree)

The average acceptance probability of the sampled configurations significantly contribute to the efficiency of, on one hand, the underlying MC algorithm, and, on the other hand the tree algorithm. Within the MC algorithm the acceptance probability is usually directly coupled with the size of the configurational changes. Increasing the size of the moves leads to decreasing acceptance rates, which decreases the number of sampled points in configurational space, but the distances between these points are enlarged. There exist a balance between a sufficiently large number of sampled points and feasible large distances between these points. Gelman *et al.* showed a theoretically optimal acceptance rate of 44% for an one dimensional problem ($d = 1$), reaching 23% for $d \to \infty$[18].

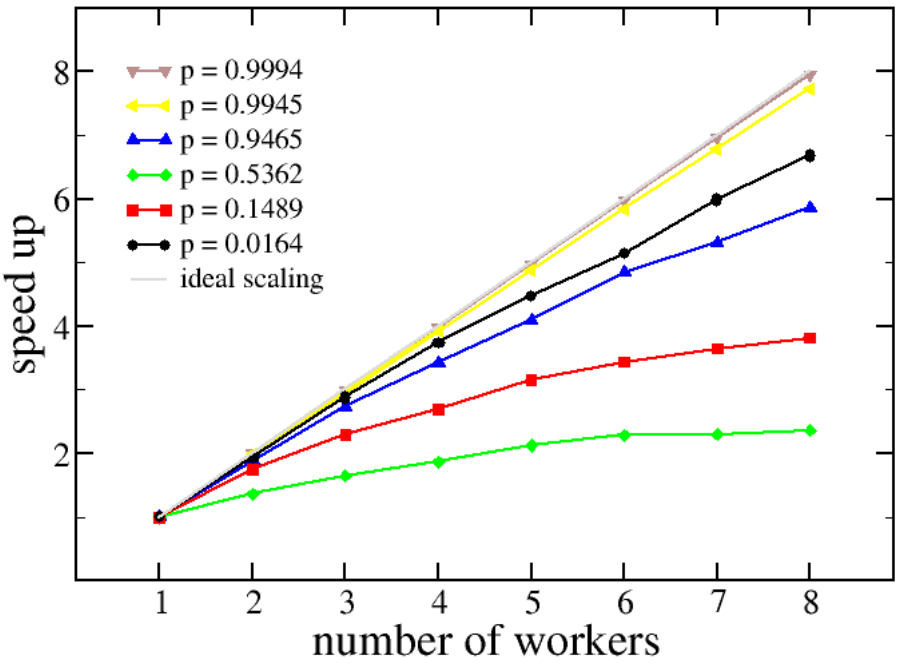


Figure 2.5: Speedup depending on the number of employed computing units for a give Monte Carlo acceptance rate. Ideally the speedup is proportional to the computing units.

On the other hand, the performance of the TMC algorithm is mainly influenced by the accuracy of the predicted Markov chain direction, decreasing the number of computed off-chain elements. Hence, TMC aims to create a slim tree in the most probable direction, as deep as possible and as wide as necessary. Thus the number of parallel computed Markov chain elements maximizes for a given number of computing units. The most probable Markov chain direction is estimated by the average and actual acceptance probabilities of the configurational changes. Theoretically, the direction can be predicted perfectly, if the acceptance would be close to zero or one. Thus, an unbalanced tree with mainly accepted or rejected branches is build, computing the elements highly parallel, see Fig. 2.5. Hence, almost all computed

elements are utilized in the Markov chain, resulting in a speedup $S \propto N_E$ (N_E is the number of computing units). Unfortunately, the sampled configurational space collapses at the initial point in both cases, since the number of accepted configurations contract to zero elements ($acc = 0\%$) or the move size is negligible ($acc = 100\%$). In contrast to the perfectly predicted Markov chain described above, an average acceptance probability of 50% span the variety of all possible Markov chain, which are all equally probable. Thus, the speedup reduce to $S \propto \log(N_E)$. Recapitulating, the MC step sizes should be tuned, leading to acceptance rates that optimally balance the tree efficiency (0% or 100% acceptance) and the MC efficiency (23% acceptance).

As a result, new binary tree elements can be created in the most probable direction of the "correct" Markov chain. Therefore, for each tree element θ_i (exhibiting an acceptance probability p_{θ_i}), the accessibility of the child elements are calculated, p_{θ_i} (acceptance branch) and $1 - p_{\theta_i}$ (rejection branch). Employing the acceptance check, the accessibilities change to 1.0 (acceptance branch) and 0.0 (rejection branch). Assuming an element θ_k is connected by m acceptance and n rejection branches to the last known Markov chain element. Its accessibility can be calculated using an average acceptance probability $\bar{p}$ by

$$\alpha_{\theta_k} = \bar{p}^m (1 - \bar{p})^n. \tag{2.10}$$

The accuracy of the estimated Markov chain direction can be improved determining the acceptance probabilities of each single element more accurately, as discussed in Sec. 3.3.1. The more precise the estimations are the more slim the tree can be sampled, reducing significantly the overall run time, shown in Sec. 4.3.

2.3.6 Optimized configurational change

The number of Markov chain elements, necessary to reach a desired accuracy in the sampled system properties, depends on the size of the configurational change and the acceptance rate. Any move is possible, which conserves (at least) the balance condition and ergodicity, see Sec. 2.1. Basically, all atoms

can be placed randomly in the simulation box within a move. However, considering complex chemical systems, this would lead mostly to physical meaningless systems, resulting in unsatisfactorily low acceptance rates. Therefore, configurational changes are selected, modifying the last (accepted) configuration, utilizing e.g. single particle displacements. Hence, the Markov chain is comparable to a random walk in configurational space. Depending on the simulated system, points of interest can be separated by large distances and high energy barriers. The move sizes can be enlarged by increasing the displacement of particles or the number of displaced particles. However, at the same time, the acceptance ratio is decreased drastically in conventional trial moves.

Optimized moves decorrelate neighboring configurations, thereby spreading the sampled configurational space, while maintaining reasonable large acceptance rates. System specific information can be utilized, e.g. molecule information (displacing whole molecules instead of atoms or rotating parts of long chain molecules etc.). Hybrid Monte Carlo[19], utilize a sequence of molecular dynamic steps, which can be based on a approximate potential[20]. Equivalently, a sub-chain of MC moves can be applied based on a approximate potential, described in Sec. 3.4.1. Additionally, unphysical moves are possible, e.g. swapping of two atoms of different kinds, which could be improbable in molecular dynamics (requiring the overcoming of large energy barriers). Furthermore, the system can be sampled at multiple temperatures and configurations can be exchanged between these temperatures. Thus the sampling of highly probable points in configurational space, separated by energy barriers, can be enhanced (see Sec. 3.4.3). Furthermore, moves can be specialized with respect to desired system properties. A good example is the proton reordering move (described in Sec. 3.4.4), improving the sampling of dipole moments. Disordered protons in a lattice structure of water ice are collectively rotated satisfying the required ice rules, which lead to significant changes in the dipole moments. Implemented specialized moves, especially for H_2O systems (liquid and solid), are treated in more details in Sec. 3.4.

In conclusion, the TMC algorithm is able to calculate most probable Markov chain elements in parallel, which decrease the overall simulation time

with at least a logarithmic scaling with respect to the number of employed computing units. Specialized moves improve the convergence of the investigated system properties. In the following, an implementation for molecular simulation is described in detail. Specialized moves, pre-sampling and accurate estimations of the Markov chain direction are described, leading to performance improvements, which is shown afterwards in Sec. 4.3.

Chapter 3

Implementation

The Tree Monte Carlo (TMC) algorithm generates configurations of possible Markov chain elements, stored in a binary tree structure. These tree elements are dynamically attached at the most probable position of the a priori unknown Markov chain. The estimations will be described in Sec. 3.3. Configurational changes and energies have to be calculated and distributed to distinct computing units. The parallel processing of the mentioned tasks reduces the overall simulation time. In this section the implementation of the TMC algorithm is described including the already mentioned optimizations.

The complete TMC algorithm is implemented in a development version of `CP2K` [21] using Fortran 95. Separated modules are implemented for TMC related tasks, including tree element handling (creation, deallocation and search routines), configurational changes, acceptance probability handling and estimations, message creation and reading, as well as analyzing the Markov chain elements. The `CP2K` environment provides features for energy calculations, communication, input file reading, physical and mathematical constants, error handling etc., which are employed in the present algorithm and not described in more detail here.

The TMC implementation is optimized for computational expensive energy calculations using e.g. DFT or ab initio methods, which usually include iterations of various expenses, e.g. self-consistent field iterations. The implemented master-worker scheme ensures continuous processing without syn-

chronization times. The master manages the tree nodes, including creation, deallocation, acceptance checks, probability updates and cancellations, described later in detail. For each available worker, a new task is created by searching the most probable tree element to proceed. Average and actual acceptance probabilities are taken into account to search this element (see Sec. 3.3). If the tree element is not already existing, it will be created first. Since the estimations of the Markov chain directions significantly influence the TMC efficiency, the first priority of the master routine is to receive the results of the workers. Thus acceptance checks can be performed, probabilities updated, and the available computing power shifted to the most likely Markov chain elements. The master procedure is outlined in Fig. 3.1, where two types of available worker groups are distinguished (described below). In general, the computational effort for energy calculations dominates the simulation. Nevertheless, the computational cost for the configurational change vary a lot with the type of move selected. For example, a single particle move (cheap) can be easily performed on the master side, but Hybrid Monte Carlo or Nested Monte Carlo requires a huge number of calculations, which could easily overload the capacity of the master unit. Hence, these expensive moves can be transferred to a worker. However, as the scaling in parallelization of the configurational changes and energy calculations can differ significantly, this requires worker groups of different sizes. Consequently, the user can select one of the following setups:

- configurational changes (cheap) on master unit, energy calculations on workers,
- configurational changes and energy calculations on the same worker (serial, no pre-sampling)
- configurational changes (expensive) and energy calculations on distinct workers; separated types with different sizes.

Additionally, a group for analyzing the resulting Markov chain on the fly can be selected by the user, which is not described in detail here. Hence, the available processors (without the master unit) are redistributed into three

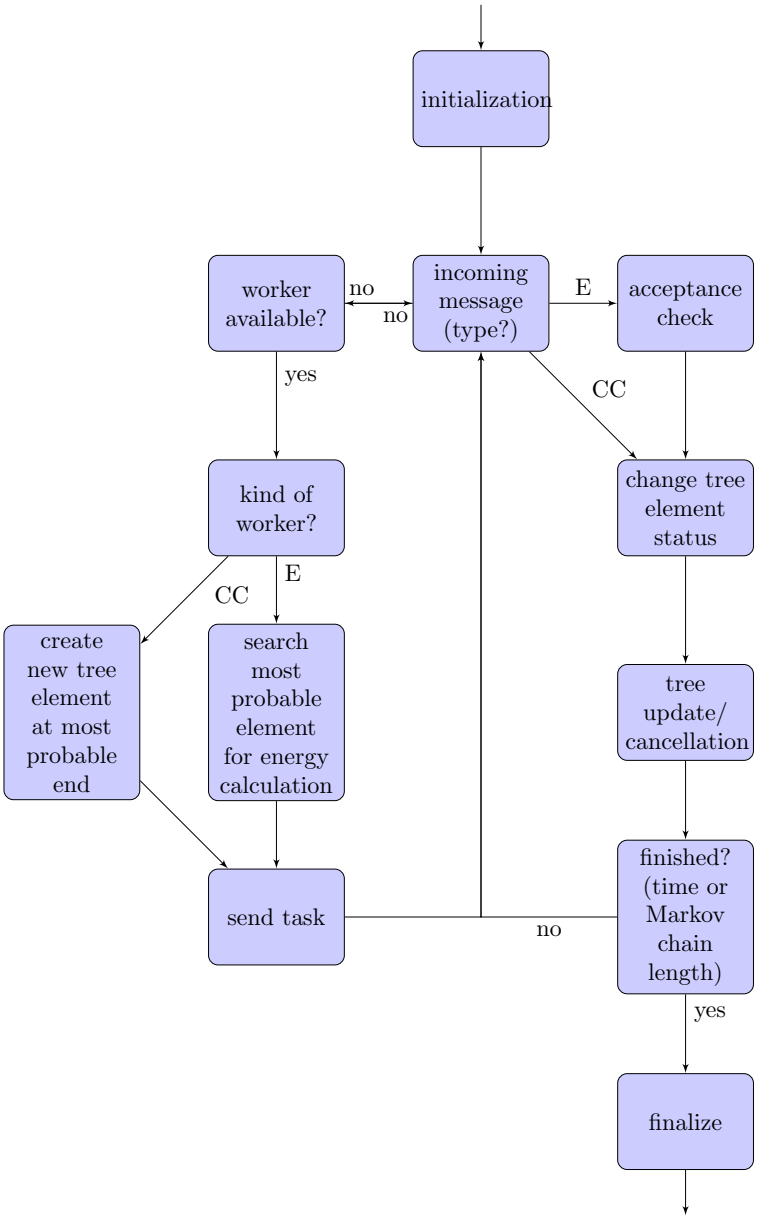


Figure 3.1: The master working scheme, handling task and messages for worker groups, configurational change (CC) and energy (E).

possible types of worker groups of various sizes: groups for energy calculations, groups for configurational changes, and an analysis group. A group leader is present in each worker group, communicating with the master, dis-

tributing the received task to the group members, and querying for messages frequently, which enables an efficient dynamic cancellation of jobs. Via the

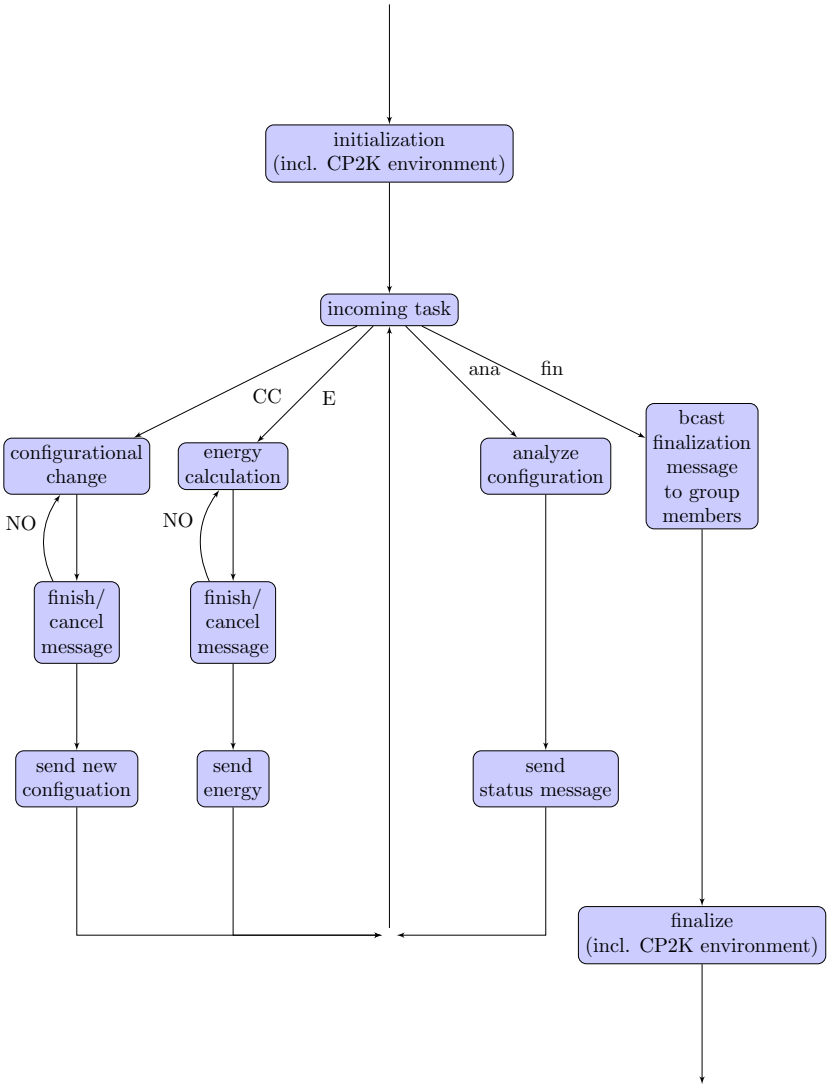


Figure 3.2: Worker group communication scheme, handling different types of messages: configurational change (CC), energy (E), analysis (ana) and finalizing (fin). The configurational change as well as the energy calculation tasks are distributed within the group and processed together. In case of cancellation, the result message only contains a flag for receipting the cancellation

group leaders, the groups are completely controlled by messages of the master routine, sketched in Fig. 3.2.

3.1 Parallelization

The algorithm is implemented in a two level parallelization scheme utilizing the Message Passing Interface (`MPI`) [10] for communication between the distinct processes (Fig. 3.3). The first level of parallelization is implemented in

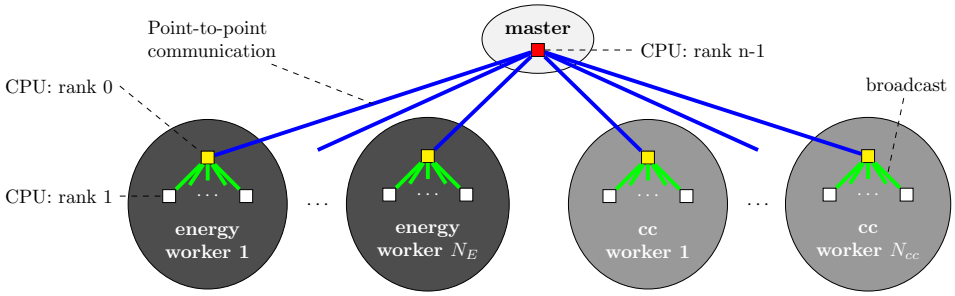


Figure 3.3: The two level master-worker parallelization scheme, including two distinct kind of worker groups for energy calculations and configurational change. The master (red) - group leader (yellow) communication is implemented in point-to-point communication (blue lines), while the communication within the groups employ broadcast routines (green lines).

the mentioned master-worker scheme, the second parallelization collects the processes into groups, mainly to distribute the sampled system, employing domain decomposition methods. For each group a group leader is defined, communicating (using point-to-point communication) with the (global) master. Hence, the master process sends the upcoming tasks to the group leader, which distributes the message to the group members using collective communication (broadcast). The result messages are handled the opposite way back to the master process. Furthermore, the working groups can employ hybrid `MPI/openMP`[22], pure openMP frameworks[23] or additional accelerators (GPUs).

The total number of available processors (N_{tot}) is divided into a predefined number of groups. The group sizes (s_E) and the number (N_E) of workers

for energy calculation determine the parallel speedup, through the average number of parallel computed Markov chain elements (within a certain run time) and the reduction in energy calculation time (with respect to domain decomposition). Furthermore, the group sizes (s_{cc}) and the number (N_{cc}) of worker groups for configurational changes should be large enough to guarantee the preparation of configurations, avoiding delays in energy calculations. It should be mentioned that the number of processors (s_E and s_{cc}) is constant within the group types and should be selected based on the scalability of the investigated functionals and system. Additionally, one analyzing worker can be employed, due to computationally inexpensive calculations using a single core ($N_{ana} = 1$). This worker could be easily parallelized (distributing configurations to separate processors), as the computation expenses of analysis tasks increases or the frequency of task overload the capacity of this worker. Since its computational expenses are low - its purpose being mainly to organise the work flow by creating tasks and handling messages - the master also consists of one single core ($s_M = 1$).

The available processors are redistributed to the related groups in a way that the energy groups fit best on the existing computing node structure[1]. Practically, computing nodes are allocated completely, which not compulsory fit with the required number of processes, as a result, idle processes s_{idle} remain. Accordingly, the total number of allocated processors sum up to

$$N_{tot} = N_E \cdot s_E + N_{cc} \cdot s_{cc} + N_{ana} + s_{idle} + s_M. \tag{3.1}$$

Summarizing, an increasing number of processors can be distributed in various setups, which effects the calculation times as follows (neglecting idle processors):

- Minimum, two processors (one master, one worker) are necessary to run the program, which reduces the algorithm to a standard serial Monte Carlo program with additional communication.

[1]Massive parallel computers are usually structured in computing nodes, where communication within the node is faster than among the nodes. Furthermore often the communication delay increases with the node distance.

- One large worker (except for the master all processors belong to one group), the computed system is distributed within the group, utilizing the domain decomposition method. As a result the computing time of single tasks reduces.
- Multiple energy groups (processors evenly distributed to the groups), multiple predicted configurations are computed in parallel (including the configurational change). The average time for computing a Markov chain element is reduced.
- Groups for configurational change and groups for energy calculations, the pre-sampling of configurations ensures continuous energy calculations (especially when using the Nested Monte Carlo method (see Sec. 3.4.1). The time for computing the configurational change reduce to a negligible value and the parallel energy calculations reduces the average time for computing the Markov chain elements.

As a result, Markov chains are sample which are equal to a serial simulation, but in a reduced run time. Additionally, multiple TMC runs can be processed in parallel (called farming within the `CP2K` environment), which is comparable to a TMC version of the standard parallel Mont Carlo method. Thus, multiple Markov chains are sampled in parallel, where the performance of the sampling of each Markov chain is improved by the speculative approach of TMC. As a result, all elements of all sampled Markov chains are averaged. Independent of the computational setups, the same Markov chain is created (equal for employing the same random number seed, described in Sec. 4.1), but the run time can significantly vary with the setup.

3.2 Tree element handling

The TMC algorithm dynamically creates binary tree(s), consisting of possible Markov chain elements, which are calculated in parallel. Each tree is a double-linked list of elements, mapping the relations between the configurations and permitting the search up- and downwards (back and forward in the Markov chain). The tree elements contain, beside the configurational data,

pointers to the parent element and two possible children (assuming acceptance and rejection of the actual element), sketched on the right side of Fig. 3.4. Assumptions (acceptance/rejection) are easily selected by comparing pointers, necessary for a dynamic movement within the tree. Furthermore, the tree elements also include status flags, tree node numbers, random seeds, move type, etc., which are used to maintain consistency and the possibility of debugging.

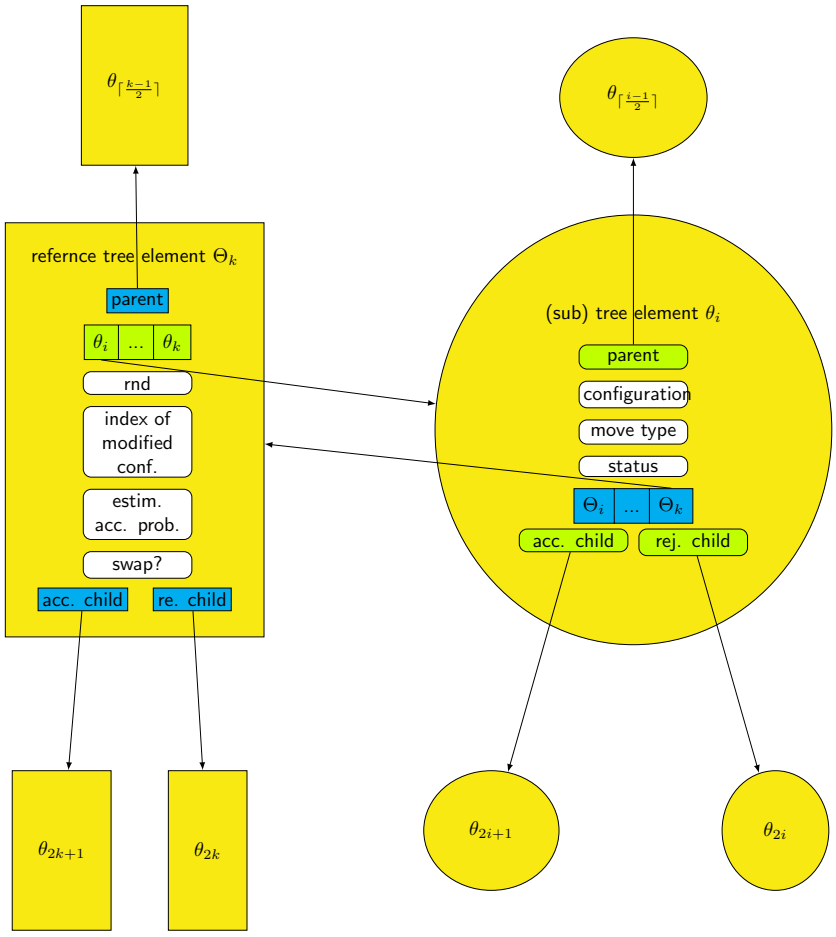


Figure 3.4: Global tree node (rectangular) and (sub-)tree node (round) including the important stored data. Pointer references are displayed with arrows, connecting parental and child elements. Furthermore, a list of pointers point to the related (sub-)tree elements respectively to all referring global tree elements, connecting between the trees.

At the beginning of its life time, configurational data are copied from the last assumed accepted element, which is found by searching up the tree (described in Sec. 2.3.1). The configurational change is then performed (if necessary on a distinct computing unit), based on a move type which is randomly selected from a user-specified list. This modified configuration is sent to an available worker and the energy calculated. If provided, intermediate energy information is received form the worker, stored in the element, and used to estimate the final energy and the acceptance probability. In the meantime, additional tree elements can be created, but not before the configurational change of the related last assumed accepted element is finished. Hence, new elements can be always attached on the rejected branch of an actual element (even if the configurational change is being processed) because these new elements are based on previous elements. Additionally, the actual status of the tree element life is recorded in a flag, which is continuously updated. At the end of its life time, unnecessary elements (elements outside the chain) are detached from bottom to top, avoiding holes in the tree and thus dangling pointers. The trajectory elements, which are not needed anymore, are detached afterwards from top to bottom (for elements without children beside the chain) up to the most recent accepted element. Thus, only a small part of the tree is kept in memory, mainly consisting of elements being processed by the workers and the necessary last accepted element.

An additional global tree is created to provide the possibility of performing parallel tempering (described in Sec. 3.4.3). This tree manages information of configurational exchanges (swap moves) between temperatures. Similar to (sub-)tree elements, pointers to parent and child elements are stored, as well as data to maintain the consistency of the algorithm and for debugging, including a node number, a random number for acceptance checks, and an estimated acceptance probability (left side of Fig. 3.4). However, the main key is an array of pointers, pointing to (sub-)tree elements, one for each temperature. Positions in the array determine the utilized temperatures, hence, a swap move is simply performed by an exchange of pointers. Assuming both, acceptance as well as rejection of swaps, a configuration can be possibly employed in Markov chains of two temperatures. Consequently,

the children of these possibly swapped elements can be employed in both temperatures as well. Since these elements are handled globally and exist uniquely in the (sub-)trees, double calculations are avoided. Therefore, configurational changes of these children are independent of the finally employed temperature.

A new task is created for the most probable global tree element, which is determined by testing the status and acceptance probabilities of previous elements, starting from the last trajectory element (finally accepted or rejected). At the creation time, the pointer array is copied from the parent element, the unique random number is calculated, and the estimated acceptance probability is initialized. Applying the configurational change, either two pointers are swapped or a single configuration is modified, selecting one temperature. In the latter case, the child of the related (sub-)tree element is selected, taking the assumptions (acceptance/rejection) of the global tree into account. If not already existing, the (sub-)tree element has to be created first as described above. Thus, the assumptions of the global tree are reflected in the (sub-)trees structure. In case of single temperature calculations, the global tree structure is equivalent to the (sub-)tree structure, with 1:1 pointer mapping. However, in case of multiple temperatures and applied swaps, the mapping becomes more complex. The assumption of the certain (sub-)tree elements are related to the assumption of the global tree element, which last modified the this (sub-)tree. Since (sub-)tree elements can be employed multiple times, the acceptance checks are applied for the global tree elements, taking the energies of either the two exchanged (sub-) tree elements or the modified (sub-)tree element and its last assumed accepted element in the related (sub-)tree into account. The employed random number is stored in the global tree element, because it is additionally utilized to calculate the acceptance probability estimations using intermediate energies. Receiving a new intermediate energy, the estimated acceptance probability has to be updated for all global tree elements, which employ this configuration. For this reason, each (sub-)tree element contains a dynamic list of pointers, pointing to all related global tree elements, simplifying the search. Thus, new global tree elements are added to the list of the related

(sub-)tree element at its creation time and erased respectively when deallocated. As a result of a performed acceptance check or acceptance probability estimation, the accessibility of tree elements changes, especially for elements on the rejected/accepted branches of accepted/rejected elements. If they are considered as not necessary any more, elements can be canceled (during calculation) or deleted (marked first). Indeed, global tree elements can be delete immediately (removing the reference pointers in the related (sub-)tree elements), while (sub-)tree elements can only be deleted, if no pointer to a global tree element is existing anymore.

As a consequence of the described tree element handling, the Markov chain elements are distributed over the created (sub-)trees, as configurational exchanges are applied. Step by step, the global tree trajectory elements are analyzed, integrating the related (sub-)tree elements into the Markov chain of the related temperature, which is specified by index of the pointer array. The configurational data of accepted Markov chain elements are stored in a file and (if requested by the user) send to the analysis worker. The number of following rejected configuration (at a given temperature) determines the weight of the previous accepted configuration.

3.3 Markov chain estimations

As already discussed, the TMC algorithm is theoretically suitable to simulate the whole Markov chain within the time of computing one configuration, neglecting the time for configurational changes and communication as well as assuming an infinite number of processors. In practice, the number of processors is limited and thus the number of simultaneous computed Markov chain elements. Indeed, elements are calculated on the Markov chain and beside the chain, which are necessary for the parallel implementation, but discarded for averaging the system properties. The ratio of the number of both types of elements determine the parallel speedup. Hence, the TMC algorithm is optimized concentrating the available computing power on the a priori unknown Markov chain. Theoretically, each way through the binary tree represents a possible Markov chain, but the probabilities resulting in

the "correct" Markov chain vary a lot, depending on the energies and the utilized random numbers.

The algorithm creates a slim tree around these "correct" Markov chain by estimating the most probable positions for attaching new elements. Therefore, the (average and actual) acceptance probabilities of the applied moves are taken into account, calculating the element accessibility. Trivially, average acceptance probabilities (eq. 2.10) can be employed, distinguishing different move types and temperatures. The accuracy is improved, estimating the actual acceptance of each tree element separately, using additional information, which will be described in more detail in Sec. 3.3.1. These values are employed to calculate the accessibility α_{θ_k} for each element θ_k

$$\alpha_{\theta_k} = \begin{cases} \alpha_{\theta_l} p_{\theta_l}, & k = 2n \\ \alpha_{\theta_l}(1 - p_{\theta_l}), & k = 2n+1 \end{cases} \quad n \in \mathbb{N}, \tag{3.2}$$

using the (estimated) acceptance probability p_{θ_l} and the accessibility α_{θ_l} of the parent element θ_l ($l = \lfloor \frac{k}{2} \rfloor$). Thus, the accessibility of already created and new elements changes when obtaining additional information of previous elements. Clearly, a performed acceptance check results in an acceptance 1.0 or 0.0 for acceptance or rejection of an element. As a result, all elements in an acceptance branch of a rejected element result in an accessibility of 0.0 and can be discarded, the same for the opposite case. Thus, calculations of these unnecessary configurations can be canceled, and is shifted the computing power to more probable elements. This dynamic process canceling is described later in Sec. 3.3.2. Therefore, accurate estimate of the acceptance probabilities improve the performance significantly.

3.3.1 Acceptance probability estimation

As an initial guess for the acceptance probabilities, an user defined value p_0 is employed, assuming that the move sizes are tuned with respect to the MC efficiency. During the simulation, average vales $\bar{p}$ are obtained, utilizing information of already performed acceptance checks of Markov chain elements, which are simply the ratio between the number of accepted and rejected

elements

$$\bar{p} = \frac{N_{\text{acc}}}{N_{\text{acc}} + N_{\text{rej}}}.$$

These values are computed and stored for each move type and each temperature separately.

Additionally, the actual values p_{θ_i} can be estimated more precisely employing intermediate energy information during the calculation. It is assumed that the energy calculations, mostly DFT and wavefunction-based methods, require iterations of self-consistent field (SCF) equations, leading to the minimum energy. Final energies are extrapolated, assuming a strictly monotonic (exponential) decreasing function. The last three intermediate energies of the SCF steps (U_i, U_{i-1}, U_{i-2}) are employed in a three point extrapolation, deploying a system of equation,

$$\begin{aligned} \exp(3a' + b') + c' &= U_i, \\ \exp(a' + b') + c' &= U_{i-2}, \\ \exp(2a' + b') + c' &= U_{i-1}, \end{aligned} \tag{3.3}$$

ignoring degenerated cases (two intermediate energies are equal). This can be rewritten as

$$\begin{aligned} a^3 b + c &= U_i, \\ ab + c &= U_{i-2}, \\ a^2 b + c &= U_{i-1}, \end{aligned} \tag{3.4}$$

which is solved for a, b and c,

$$\begin{aligned} a &= \tfrac{-U_i + U_{i-1}}{U_{i-2} - U_{i-1}}, \\ b &= \tfrac{(U_{i-2} - U_{i-1})^3}{(-U_i + U_{i-1})(U_i - 2U_{i-1} + U_{i-2})}, \\ c &= \tfrac{-U_{i-1}^2 + U_{i-2}U_i}{U_i - 2U_{i-1} + U_{i-2}}. \end{aligned} \tag{3.5}$$

Hence, the final energy can be estimated as

$$U_f = a^7 b + c \tag{3.6}$$

This estimation includes an uncertainty, thus the correct energy (U) is ex-

pected to be Gaussian distributed around this extrapolated value $U_f = \mu$,

$$p_i = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(U-\mu)^2}{\sigma^2}} \tag{3.7}$$

where the variance σ is determined by the difference to the last observed intermediate energy ($\sigma = U_i - U_f$). Furthermore, the accuracy and stability of the procedure are improved, employing two estimations $U''_f(U_{i-1}, U_{i-2}, U_{i-3})$ and $U'_f(U_i, U_{i-1}, U_{i-2})$, combining both resulting distributions to one Gaussian (eq. 3.7), where $\mu = U''_f$ and $\sigma = \max(\sigma'', |U'_f - U''_f|)$. Finally, the tree element acceptance probability p_{θ_i} is estimated by the overlap of the computed energy distributions (probability of energy differences) of the last accepted (θ_a) and the actual (θ_i) element,

$$p_{\theta_i} = \int_{-\infty}^{\infty} p_j(\theta_a) p_k(\theta_i) dU, \tag{3.8}$$

where j and k are the number of obtained intermediate energies of the related configuration. More elegant, the two energy probability distributions ($p(\theta_a)$ and $p(\theta_i)$) can be combined to another Gaussian with the mean and variance

$$\begin{aligned} \mu &= \frac{(\sigma^2(\theta_a) + \sigma^2(\theta_i))\bar{\mu} + (\mu(\theta_i) - \mu(\theta_a) - \mu_c)\bar{\sigma}^2}{\sigma^2(\theta_a) + \sigma^2(\theta_i) + \bar{\sigma}^2} \\ \sigma &= \frac{\sqrt{\sigma^2(\theta_a) + \sigma^2(\theta_i)}\bar{\sigma}}{\sqrt{\sigma^2(\theta_a) + \sigma^2(\theta_i) + \bar{\sigma}^2}}, \end{aligned} \tag{3.9}$$

employing additional information of already checked elements: the average difference of the total energy $\bar{\mu} = \langle (U_n - U_o) + (U^*_n - U^*_o) \rangle$, its uncertainty $\bar{\sigma} = \sqrt{\langle [(U_n - U_o) + (U^*_n - U^*_o)]^2 \rangle - \bar{\mu}^2}$, as well as the actual difference of the classical energies $\mu_c = U^*(\theta_i) - U^*(\theta_a)$ (when used approximate potential methods). Hence, acceptance probability can be easily calculated utilizing the error function and the (stored) random number n_r,

$$p_d = \frac{1}{\sqrt{2}\pi\sigma} \int_{-\infty}^{\infty} e^{-\frac{1}{2}\frac{((-\ln(n_r)/\beta)-\mu)^2}{\sigma^2}} = 1 - \tfrac{1}{2}\mathrm{erfc}\left(\frac{(-\ln(n_r)/\beta) - \mu}{\sigma\sqrt{2}}\right). \tag{3.10}$$

Utilizing accurate acceptance probability estimations, the performance can be significantly increased, especially for calculations with a large number of iterations ($\gg 4$). Thus, the calculations of unnecessary elements can be canceled at an early stage.

3.3.2 Dynamic cancellation

The speculative approach of TMC implicates calculations of elements off the resulting Markov chain, which are discarded for averaging the system properties. On one hand, elements in the accepted branch of a rejected element (and vice versa) are assessed to be (definitely) unnecessary. On the other hand, the accessibility of elements is influenced dynamically by the acceptance estimations, as intermediate energy information is obtained. The uncertainty of these estimations decreases with an increasing number of intermediate energies. Thus, the most probable direction of the "correct" Markov chain varies during the simulation time. Elements first expected to be on the chain can decrease in accessibility and thus change to unnecessary elements (off the chain). If the uncertainty of the estimations are low enough, calculations on these (assumed) unnecessary elements can be canceled immediately, shifting the computing power automatically to more probable (new) elements. In these cases, a status message is sent to the responsible worker group leader, which probes regularly for incoming messages. When the message is received on the worker side, it is communicated within the group and the calculation is aborted. The cancellation is finalized with a receipt message and sent by the group leader back to the master and the group is ready to receive a new task. Hence, the algorithm efficiency is increased, minimizing the calculation time on unused configurations. In contrast to definitely unnecessary elements, which can be deallocated, tree elements not definitely off the chain are kept in memory, offering the opportunity to reactivate the calculation process, while elements definitely off the chain can be deleted.

3.4 Specialized moves

As already mentioned the performance of the simulation can be improved utilizing the parallelization approach of TMC, decreasing the average computing time of Markov chain elements. As a result, the investigated system properties are determined by an average value of a sufficiently long Markov chain, providing an estimate of the ensemble-average value of the quantity. The sampled configurational space consists of states of high probability, which can be separated widely and additionally by large energy barriers. Since the MC algorithm produces a random walk in configurational space, the convergence of the investigated system properties can be improved by decorrelating configurations and concentrating the sampling on the different high probable states. Thus, the number of elements necessary to provide an accurate estimation of the investigated system properties can be reduced, and hence the performance of the simulation increased.

Conventional configurational changes, including atom and molecule translation as well as molecule rotation, can be increased by moving a larger number of particles within each step or enlarging the size of the move. However, these increased move sizes result in increased energy differences, which drastically reduce the MC acceptance (eq. 2.5) and, as a result, many configurations are calculated but only a few employed in the chain. Hence, a compromise has to be employed between large step sizes (decorrelating configurations) and reasonable acceptance rates (number of sampled points).

Specialized or "smart" Monte Carlo moves try to decorrelate neighboring configurations, utilizing known system information, while maintaining sufficiently large acceptance rates. A variety of specialized moves exist for different kind of MC problems, which are adapted with respect to methods and/or systems. In the following sections moves are described which are adapted to accurate chemical simulations, especially for the presented liquid water and ice simulations. Assuming computational intensive energy calculations and the possibility of creating a computational cheap approximate potential, the Nested Monte Carlo method (described in Sec. 3.4.1) performs a series of MC steps based on this potential, which increases the overall con-

figurational change. Constant pressure calculations require a volume move, which is described in Sec. 3.4.2. Furthermore, the sampling of systems with large energy barriers can be improved by employing the Parallel Tempering method (see Sec. 3.4.3). Finally, the dipole moment sampling of e.g. ice Ih, is enhanced by collective molecule moves with respect to the underlying lattice structure, described in Sec. 3.4.4. Additional moves can be easily integrated in this TMC implementation.

With an increasing complexity of the configurational changes, the computational expenses could increase, which could be compensated using the pre-sampling method. However, to maintain continuous energy calculations, the time for an configurational change has to be shorter than the time for an energy calculation, independent of the number of worker groups (for configurational change). This has to be ensured because, assuming the calculation of a configuration θ_i, new elements in accepted direction (θ_{2i}, θ_{4i}, θ_{8i}, etc.) can not be created before the previous element is finally created. If the time for computing a configurational change is longer than the time for computing an energy, a delay will be introduced in the sequence of energy calculations. This delay can not be compensated by an increasing number of configurational change worker groups.

The presented moves need no velocity or force evaluation, thus the Hamiltonian ($\mathcal{H}(\mathbf{r}, \mathbf{p})$) is reduced to the potential energy $U(\mathbf{r})$ for simplicity in further notations of the acceptance criteria .

3.4.1 Nested Monte Carlo

Accurate molecular simulations employ computational expensive energy calculations, investigating e.g. DFT or ab initio methods, which dominate the computational effort. The efficiency of the algorithm is increased, reducing the number of calculated configurations, while maintaining the sampled space by spreading the points in space. The Nested Monte Carlo (NMC) algorithm reduces the correlation between Markov chain elements by creating a sub Markov chain, based on a approximate potential, retaining sufficiently large acceptance ratios[24]. Starting from an old configuration θ_o, a sequence

of (usually fixed k) MC steps is created, based on the approximate potential U^*. The final element of the NMC sequence is employed as new configuration θ_n in the exact chain, calculated with the accurate potential (U), see Fig. 3.5. Energy differences of the sampled configurations ($U(\theta_i)$ and $U^*(\theta_i)$)

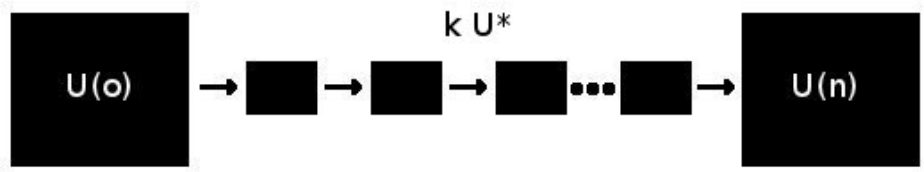


Figure 3.5: A Markov chain of length k, based on a approximate potential U^*, is used as configurational change, employing the final configuration as the new element of the exact chain. The acceptance of this element is checked utilizing the approximated and the exact energy of this configuration.

influence the MC acceptance

$$\mathrm{acc}(\theta_o \to \theta_n) = \min(1, e^{-\beta[(U(\theta_n)-U^*(\theta_n))-(U(\theta_o)-U^*(\theta_o))]}). \tag{3.11}$$

An increasing number of k NMC steps increases the distance between the sampled points in configurational space, creating configurations of high probability, supposing similar ensembles of both potentials.

Within each NMC step, any valid MC move can be performed, accepted with the usual MC criteria (Eq. 2.5) utilizing the approximate potential. Moves should be only utilized in the NMC, if the approximate potential is able to approximate the energy change for this move. For example, if the approximate potential significantly varies from the exact potential for volume changes, the volume move should be applied directly with the exact potential (not within the NMC).

The NMC method performs a significantly increased configurational change, while decreasing the computational effort for the configurational change [25, 20]. Thus, the simulations of system properties of a given variance is reduced in the number of necessary Markov chain elements. Assuming the displacement of N particles by employing N single particle moves, the original MC

algorithm requires a computational effort of

$$W_{MC} = N w_c \frac{1}{\alpha_c}, \tag{3.12}$$

spending w_c per correct energy calculation, and accept the moves with an average acceptance probability of α_c. In contrast to that, the NMC employed for an similar configurational change

$$W_{NMC} = \left(N w_a \frac{1}{\alpha_a} + \frac{N}{k} w_c \right) \frac{1}{\alpha_{ac}} \tag{3.13}$$

applying k single particle moves within a sub chain, each employed with a computational cost w_a for calculating the approximate potential, and accepted in average with an probability α_a. The whole configurational change is accepted in the exact chain with an average probability α_{ac}, employing one exact energy calculation. This acceptance, comparing the two potentials, depends on the slope of the energy difference of the exact and the approximate potential with respect to the number of NMC steps. The computational effort is reduced by a factor of

$$\frac{W_{MC}}{W_{NMC}} = \frac{\alpha_{ac}}{\left(\frac{w_a}{w_c} \frac{\alpha_c}{\alpha_a} + \frac{1}{k} \alpha_c \right)} \tag{3.14}$$

In other words, the speedup increases with the number of NMC steps, see also [26], assuming roughly same probabilities to accept the move in either the exact or the approximate potential. For example, assuming the average acceptance of $\alpha_a = \alpha_c = 20\%$ and a difference in computational effort of $w_c = 30 w_a$, employing $k = 30$ NMC steps, the speedup result in a value of $25\alpha_{ac}$, depending on the two potential acceptance probability. Despite of the low computational cost of the approximate potential, the whole configurational change can become computational expensive, especially employing a large number of NMC steps. Furthermore, the scaling with respect to the utilized number of processors could be lower compared to the exact potential. Therefore, the NMC move can be performed on separate working groups of tuned sizes, ensuring continuous energy calculations without delays, using

the speculative approach of TMC. Therefore, the number of configurational change workers has to be large enough to guarantee prepared configurations for all energy workers. However, the time for the configurational change (k NMC steps) has to be shorter than the exact energy calculation time, especially for acceptance branches, otherwise delays would be introduced in the exact energy calculations.

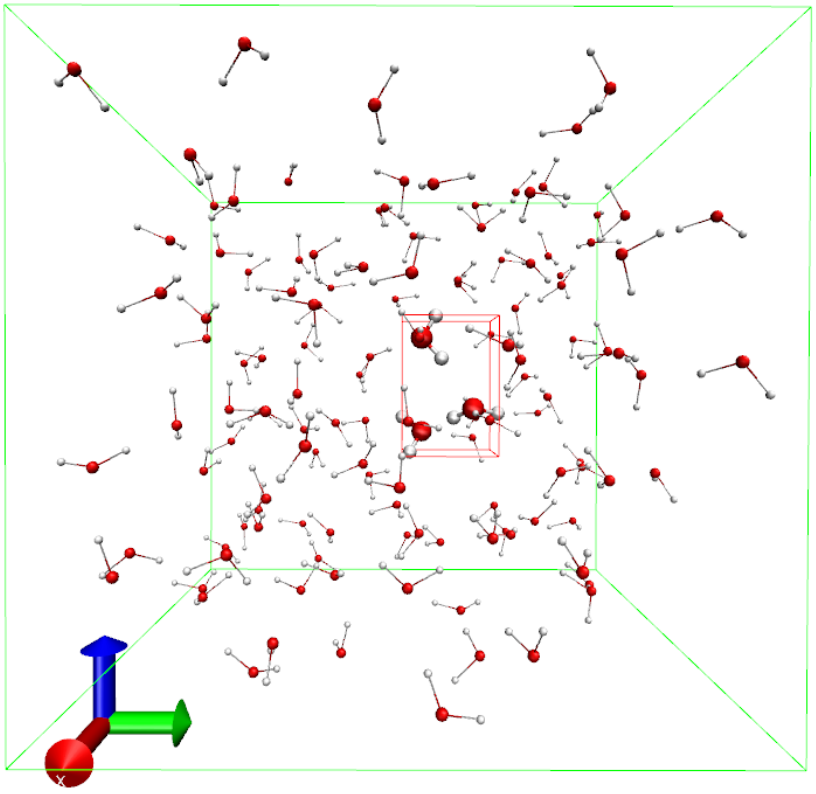


Figure 3.6: Moves applied using a sub-box modify only atoms within the sub-box (bold molecules), other atoms are fixed. The green box is the whole simulation cell and the red box is the sub-box, the center is selected randomly

As long as the approximate potential captures adequately the physics of the full potential, reasonable acceptance rates α_{ac} are maintained, which can be improved e.g. fitting the approximate potential with respect to the exact one, see Iftimie and Salahub [24]. Furthermore, the resulting energy differences can be limited by performing the moves of the sub chain only on

particles within a truncated space (called sub-box, see Fig. 3.6), selecting the sub-box center uniformly in the simulation cell, which preserve the ergodic sampling, using boundary conditions. Hence, the simulation cell is divided into randomly selected atoms/molecules $\mathbf{r}_s$ and fixed particles $\mathbf{r}_r$, where the move $\begin{pmatrix}\mathbf{r}_s\\ \mathbf{r}_r\end{pmatrix} \rightarrow \begin{pmatrix}\mathbf{r}'_s\\ \mathbf{r}_r\end{pmatrix}$ lead to regular sampling, which retain the Boltzmann distribution invariant, proved by Manousiouthakis [9]. Detailed balance is conserved, rejecting moves, which displace atoms/molecules out of the sub-box (utilizing the geometrical center). However, in case of rotational moves, atoms can successfully leave or enter the sub-box, which is a reversible move, retaining the detailed balance condition.

3.4.2 Volume change

Constant pressure (NPT) calculations require random volume changes, leading to the equilibrium volume during the simulation. In the following, the acceptance criteria is derived from the partition function, which can be written as

$$Q(N,P,T) = \frac{\beta P}{\Lambda^{3N} N!} \int dV V^N \exp(-\beta P V) \int d\mathbf{r}^N \exp(-\beta U(\mathbf{r}^N)), \quad (3.15)$$

for N particles and a defined pressure P, for more details see Ref.[8]. Applying isotropic moves in uniform distributed volume $V' = V + \Delta V$, where $\Delta V \in [-\Delta V_{max}, +\Delta V_{max}]$, the probability density can be written as

$$N_{N,P,T}(V) = \frac{V^N \exp(-\beta P V) \int d\mathbf{r}^N \exp(-\beta U(\mathbf{r}^N))}{\int_0^{V_0} dV' V'^N \exp(-\beta P V') \int d\mathbf{r}^N \exp(-\beta U(\mathbf{r}'^N))}. \quad (3.16)$$

Hence, a new configuration θ_n (including a modified volume V' and scaled coordinates of particles) is accepted with the probability

$$\mathrm{acc}(\theta_o \rightarrow \theta_n) = \min(1, e^{-\beta[(U(\theta_n)-U(\theta_o))+P(V'-V)-Nk_BT\ln(V'/V)]}). \quad (3.17)$$

The described move is mostly applied on cubic cells. One might argue that the change in volume can be also applied in one direction of the cell (L_i),

resulting in an anisotropic sampling. Hence, the cell changes with $L_i A_{jk} = V \rightarrow V + \Delta V = L_i' A_{jk}$, where $L_i' = V'/A_{jk}$ and A_{jk} is the surface spanned between the two other directions, conserving the acceptance criteria.

Furthermore, the volume move can be performed, sampling uniform distributed edges $L' = L + \Delta L$, where $\Delta L \in [-\Delta L_{max}, +\Delta L_{max}]$. Assuming isotropic moves on a cubic cell $L = V^{1/3}$, the integration variable dV in Eq. 3.15 can be substituted by $dV = 3dL\ V^{2/3}$, resulting in the acceptance criteria

$$\mathrm{acc}(\theta_o \rightarrow \theta_n) = \min(1, e^{-\beta[(U(\theta_n)-U(\theta_o))+P(V'-V)-(N+\frac{2}{3})k_B T \ln(V'/V)]}). \quad (3.18)$$

Equivalent, anisotropic moves can be performed, scaling single edges $L_i A_{jk} = V \rightarrow V' = V + \Delta V = L_i' A_{jk}$, substituting the integration variable in Eq. 3.15 $dV = dL A_{jk}$. Since A_{jk} remains constant (within the move) the acceptance probability can be again written similar to Eq. 3.17. The mentioned moves converge to equal distributions, shown for an ideal gas in Fig. 3.7.

3.4.3 Parallel Tempering

Simulating systems with large energy barriers, single temperature simulations may become trapped at local energy minima, during typical timescales of a simulation. Parallel Tempering (PT)[27, 28], or replica exchange, provides the possibility to exchange configurations of large separated points in phase space. Therefore, the system is duplicated and sampled at M different temperatures (of a sufficiently large range $[T_{min}, T_{max}]$). High temperatures systems are generally able to sample large volumes of phase space, while low temperature systems allow precise sampling in local regions. The additional configurational exchange (swap) between these systems, provide the possibility to access a large set of local regions at low temperatures. An attempted exchange of two configurations (with related inverse temperatures β_i and β_j) is accepted using

$$\mathrm{acc}((\theta_i, \theta_j) \rightarrow (\theta_j, \theta_i)) = \min(1, e^{-(\beta_i - \beta_j)(U(\theta_i) - U(\theta_j))}). \quad (3.19)$$

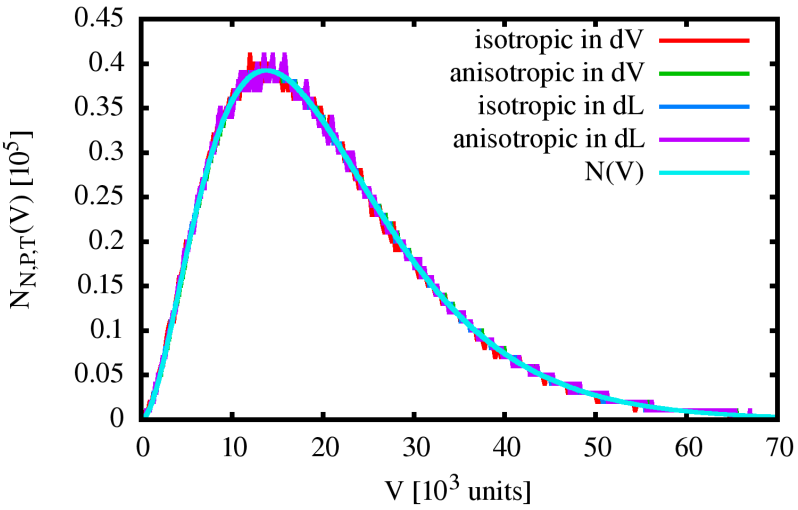


Figure 3.7: The probability density of the volume move in a ideal gas, investigating isotropic (red) and anisotropic (green) moves of fluctuating volume (dV) and isotropic (blue) and anisotropic (violet) moves of fluctuation edge L (dL). The analytic result of the sampled system at 50K, 1bar and two particles is given by Eq. 3.16.

Considering the distributions of energies of two distinct temperatures, the overlap of these distributions decrease with increasing temperature difference. Hence, typically neighboring temperatures are selected for the exchange. Furthermore, the PT technique can be also combined with molecular dynamics, which is illustrated and bench-marked by Bussi [29], but not discussed here in detail. The number of necessary replicas, the frequency of swaps and implementations using this Hybrid MC is discussed by Earl and Deem [27]. It should be mentioned that the sampling of M replicas requires M times more computational effort, but the PT is more than $1/M$ times more efficient than a standard, single temperature MC simulation. This increase derives from allowing lower temperature systems to sample regions, which would had not been reached in a M times longer single-temperature simulation. An additional benefit is the generation of results for a range of temperatures, which may also be of interest to the investigator, e.g. for finite temperature behavior of the heat capacity[30].

The implementation of the speculative approach of TMC combined with parallel tempering requires a global tree, which is described in Sec. 2.3.2). This tree allows the handling of possible outcomes of swap moves, while avoiding duplications of configurations. Therefore, a new tree element (performing a swap) first copies the pointer array from the previous element and exchanges two pointers. However, the related (sub-)tree elements can be already assumed to be rejected (in a previous step), but only accepted configurations can be swapped. Thus the acceptance check is performed with the last assumed accepted configurations of the related (sub-)trees (which are traced first, but not store in the global tree element). As a result, the elements stored in the pointer array are not (necessarily) the exchanged ones, but anyways following configurational changes are performed on basis of the last assumed accepted elements. In that way, actual positions in the (sub-) trees are preserved for the case of rejecting the swap move.

Combining the PT with the NMC moves, two configurations can be connected by a sub Markov chain created on a certain temperature. However, the final configuration can be be employed at multiple temperatures of the exact Markov chain, using parallel tempering. Thus, the temperature for creating the sub chain is selected randomly to ensures an unbiased and reproducible (see Sec. 4.1) sampling. As a result, the NMC acceptance check (eq. 3.11) is modified with respect to the temperatures of creation β_i'

$$\mathrm{acc}(\theta_o \to \theta_n) = \min(1, e^{-[(\beta_n U(\theta_n) - \beta_n'(\theta_n) U^*(\theta_n)) - (\beta_o U(\theta_o) - \beta_o'(\theta_o) U^*(\theta_o))]}). \quad (3.20)$$

3.4.4 Proton reordering

The following move is employed to improve the dipole sampling of proton disordered ice Ih structures, utilizing the structural knowledge of the system. In ice Ih, the oxygen atoms are located on a lattice with hexagonal symmetry, while the hydrogen atoms can be distributed in many ways with respect to the Bernal-Fowler[31] rules for ice. Possible realizations of the proton arrangements are separated by large energy barriers, which results usually in long dielectric relaxation times [32]. Moving from one of these realizations

to another one require a collective molecule rotation, but significant random molecular rotations violate the ice rules, which result in dramatically low acceptance, because new proton positions do not fit in the crystal structure. Therefore, a collective proton reordering[33, 34, 35, 36] was developed satisfying the ice rules, by performing the following steps:

1. pick randomly a molecule,
2. select one if its four neighbors, which could be a proton donor or a proton acceptor,
3. select a chain of neighboring molecules randomly, with the selected proton donor/acceptor property of the first neighbor, forming a closed loop (at a certain point the chain comes back to an already selected molecule),
4. rotating the molecules in the loop around the hydrogen-bond not involved in the loop, utilizing the angle between both neighboring oxygen in the loop, with respect to the actual oxygen; the first selected molecules not attending in the loop are discarded.

Thus, the relations flip between the molecules in the loop, a proton donor gets a proton acceptor and vice versa, but the number of hydrogen bonds is conserved. This move is accepted with the standard MC acceptance criteria (eq. 2.5). As a result of the reordering, the molecular dipole directions are modified, which vary the cell dipole moment. If a loop of molecules is created within the ring of the hexagonal structure, the cell dipole changes only little, because the modifications in molecular dipole directions of opposite molecules compensate. However, if the loops cross the periodic boundaries of the simulated cell (displayed in Fig. 3.8), the cell dipole changes significantly. Thus, the proton reordering remarkably improves the dipole sampling and enlarge the sampled configurational space. The importance of this will be shown in section 5.

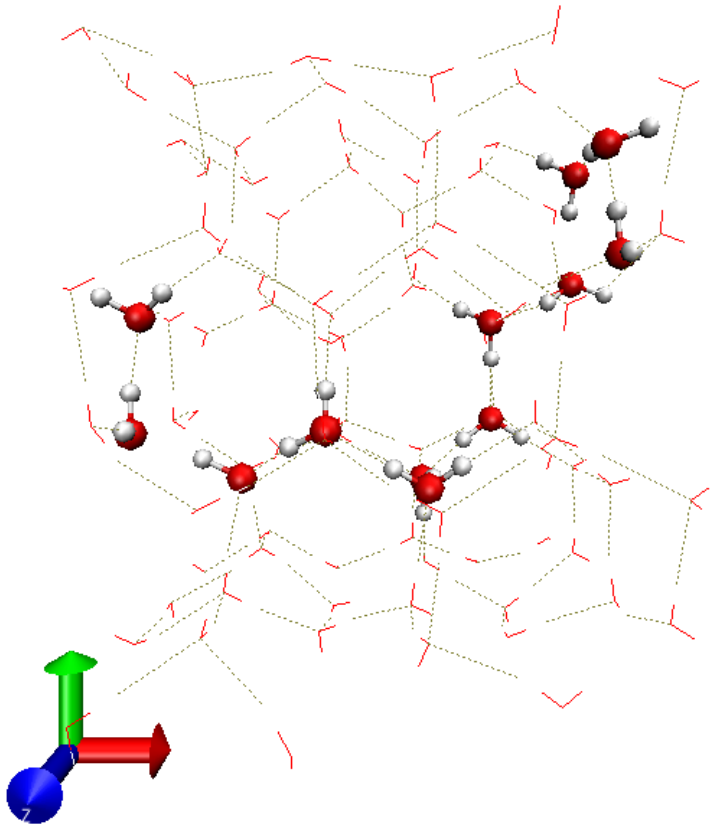


Figure 3.8: Ice Ih structure with one possible loop of molecules for proton reordering, crossing the periodic boundaries. Molecules are rotated around the hydrogen-bond not cooperating the loop, changing significantly the cell dipole moment.

3.5 Restarting

At the end of the program run time, the complete memory used is freed, including the deallocation of all tree elements and obtained acceptance probabilities. Assuming the calculation of system properties of a desired accuracy, a sufficiently long Markov chain has to be sampled. On one hand, the run time is limited on many supercomputers, which could be shorter than the necessary time for computing the desired Markov chain length. On the other hand, it could be necessary to extend simulations to obtain more accurate results. Therefore, a restart file system provide the latest status of the pro-

gram run, including the obtained information of the simulation properties (e.g. acceptance rates). At the end of each run, a restart file is written, containing the last accepted configuration(s), average acceptance probabilities, random number generator seed(s) etc. Hence, the simulation can be continued at the last point, resulting in a Markov chain equal to a longer run without interruption. Predicted elements have to be sampled again, but the already improved estimations of the Markov chain directions are preserved. Furthermore, the restart files can be additionally written regularly to restart after possible computing node failures.

Part II

Validation and Applications

Chapter 4

Validation and Performance

In this section the implemented TMC algorithm is verified, testing step by step the computational parts of a simulation starting from the random number generator up to obtained results. Furthermore, the performance of the distinct optimizations are presented using the average computing time of a Markov chain element and the resulting speedup.

Next, two applications of ice and liquid water are presented. Structural and electronic properties of the ice structures are analyzed and the phase transition temperatures from the proton disordered Ih phase to the proton ordered XI phase is estimated for hybrid and semi-local DFT functionals. Moreover, densities and radial distribution functions of liquid water under ambient conditions are calculated at the second-order Møller-Plesset perturbation theory level and compared to a few density functional approximations, including semi-local functionals, hybrid functionals, and functionals including empirical dispersion corrections.

4.1 Random number generator and tree build

A key role in every Monte Carlo algorithm is a random number generator (RNG), which is essential to conserve the balance condition and ergodicity as configurational changes and acceptance checks are performed. Furthermore, calculations have to be reproducible to provide the possibility of locating

errors in the implementation and to compare the performances of different setups calculating equal Markov chains. Developments in the `CP2K` package are tested frequently, utilizing a tool that compares results of a set of simulations covering the implemented methods. Thus, the results of these simulations have to be identical, independent of the number of employed computing units or restarting cases (and ideally the utilized computer architectures and compilers). Consequently, the possibility of a deterministic behavior of the simulation has to be provided, independent of the number of created elements beside the Markov chain.

The implemented TMC utilizes the RNG proposed by L'Ecuyer *et al.* [37], because it provides multiple random number streams (2^{64} streams of length 2^{127}), including a framework to jump from one stream to another. It can be initialized with a random seed (random case) or an user-specified seed (deterministic case). The streams and the positions in these streams are handled by the RNG seeds, which are stored in tree elements, continuously updated, and communicated to the related workers (utilizing random numbers). Thus, the sequence of random numbers can be preserved while dealing with multiple tree elements. Each tree element selects the next stream with respect to the stream of the parental element, which can be performed independent of the actual position in the stream of the parent element. Thus, reproducible Markov chains are created (and written to files), independent from the number of created tree elements and thus from the number of workers.

Furthermore, the created trees can be plotted, utilizing the graphViz [38] library. Thus, the tree creation and handling processes can be monitored and verified, including the absence of loops of pointers, correct positions, correct numbering, correct binary behavior of the tree, and deallocations (an example is shown in Fig. 4.1).

Within the development process, the Markov chain creation is compared for runs of various number of workers and worker group sizes, with and without NMC and PT, and with and without restarting (one long run and multiple short ones). As a result, identical Markov chains are produced, as the same initial random number seed is provided. In contrast to the Markov

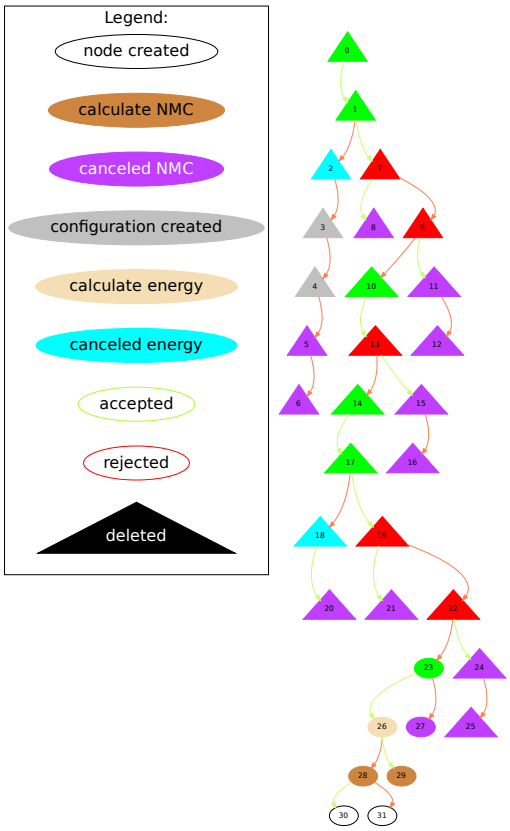


Figure 4.1: Snapshot of the created tree, with deallocated elements (triangles) and active elements (round). The different colors monitor the states of calculation from creating the tree element, creating the configurations, energy calculations, canceling and deallocation. The Markov chain elements are marked green (accepted) and red (rejected).

chains, the number of calculated elements off the chains (tree width) varies with the number of workers, as expected.

4.2 MC moves and distributions

The ergodicity and reversibility is verified monitoring the distributions of possible moves. First, it is verified that atoms and molecules are uniform

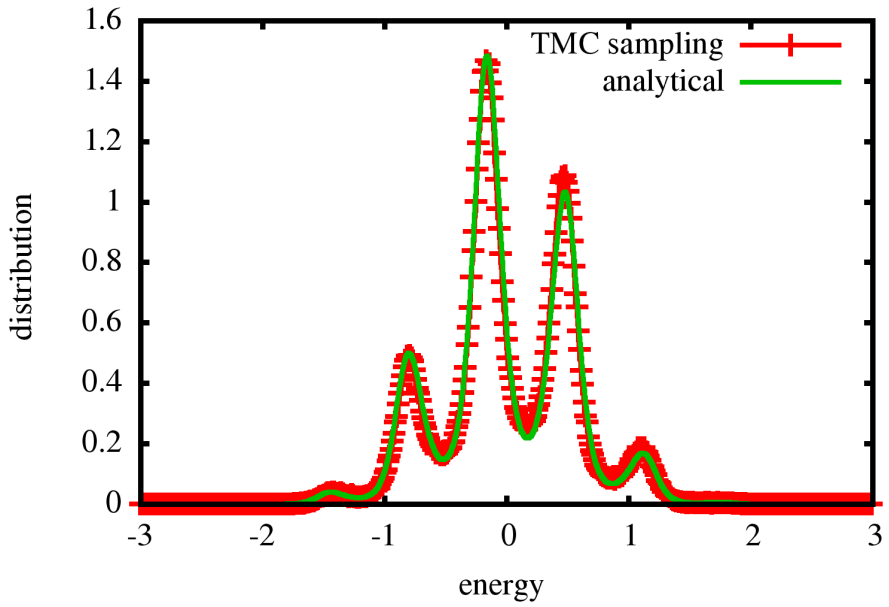


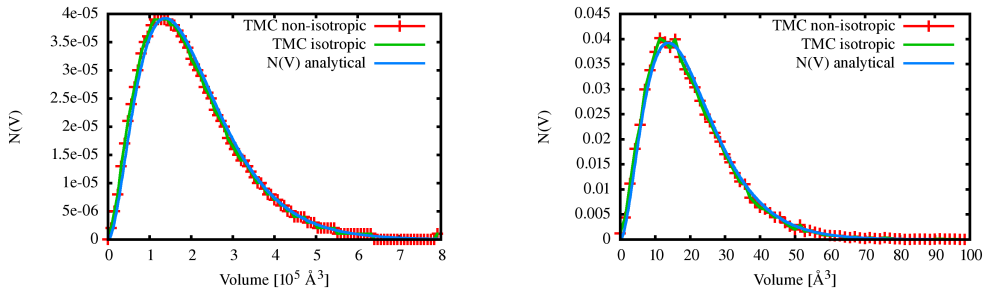
Figure 4.2: Distributions of the TMC sampled positions (x) of a single atom in an external potential (eq. 4.2), compared with the analytical result (eq. 4.1).

randomly selected, as well as the uniform distribution of the moves within the selected interval (e.g. $[\mathbf{r} - d\mathbf{x}, \mathbf{r} + d\mathbf{x}]$ for random displacements). Then, the implemented acceptance checks are verified, examining the resulting distributions and comparing them with known analytical distributions and reference calculations of the molecular dynamics package (provided by `CP2K`). In case of random displacements, the distributions of positions of an multidimensional oscillator as well as several artificial potentials are compared with the analytical values,

$$\mathcal{N}(\mathbf{r}) = \frac{\exp(-\beta U(\mathbf{r}))}{\int_{-\infty}^{\infty} \exp(-\beta U(\mathbf{r}))d\mathbf{r}}. \tag{4.1}$$

In Fig. 4.2, the results of an one dimensional case is presented, employing

$$U(\mathbf{r}) = 0.001\sin(5\mathbf{r}) + 0.0005(\mathbf{r})^2. \tag{4.2}$$



(a) 1bar (b) 1000bar

Figure 4.3: Volume distribution of a system of two particles at 50K of TMC simulations using isotropic (green) and non-isotropic (red) volume moves. The results are compared to the analytical distribution (blue) of Eq. 4.3.

Consequently, the correct handling of the calculated energies is monitored as well.

In case of NMC, the implementation is firstly verified using a single NMC step and the approximated potential equal to the exact potential ($U^* = U$, $k = 1$, see Sec. 3.4.1). As expected, the resulting Markov chain is identical to a single potential TMC simulation (using U), including acceptance rate of 100% for the two-potential checks. Then, investigating a distinct approximate potential, the simulations result in the correct properties for the exact potential (e.g. the distribution of positions of a multidimensional oscillator), independent of the number of NMC steps.

The volume move (NPT simulation) is verified by comparing the result to the analytical distribution of the volume of a ideal gas,

$$\mathcal{N}(V) = \frac{V^N exp[-\beta PV] \int d\mathbf{r}\ exp[-\beta U(\mathbf{r})]}{\int_0^{V_0} dV' V'^N exp[-\beta PV'] \int d\mathbf{r}\ exp[-\beta U(\mathbf{r})]} = \frac{V^N exp[-\beta PV]}{\int_0^{V_0} dV' V'^N exp[-\beta PV']}, \tag{4.3}$$

where the potential energy $U(\mathbf{r})$ is neglected for the N particles, shown in Fig. 4.3.

Finally, the TMC simulations of systems employing computationally cheap potentials are compared to molecular dynamics simulations, investigating the resulting radial distribution functions. Thus, the creation of the cor-

rect ensemble is verified, as well as the interplay of the whole computational framework, and the calculations of the analysis tool. In Fig. 4.4 the results are presented for NVT simulations of a 96 H_2O system at 295K employing flexible TIP3P potentials on each pair of atoms. The distributions coincide within the statistical error.

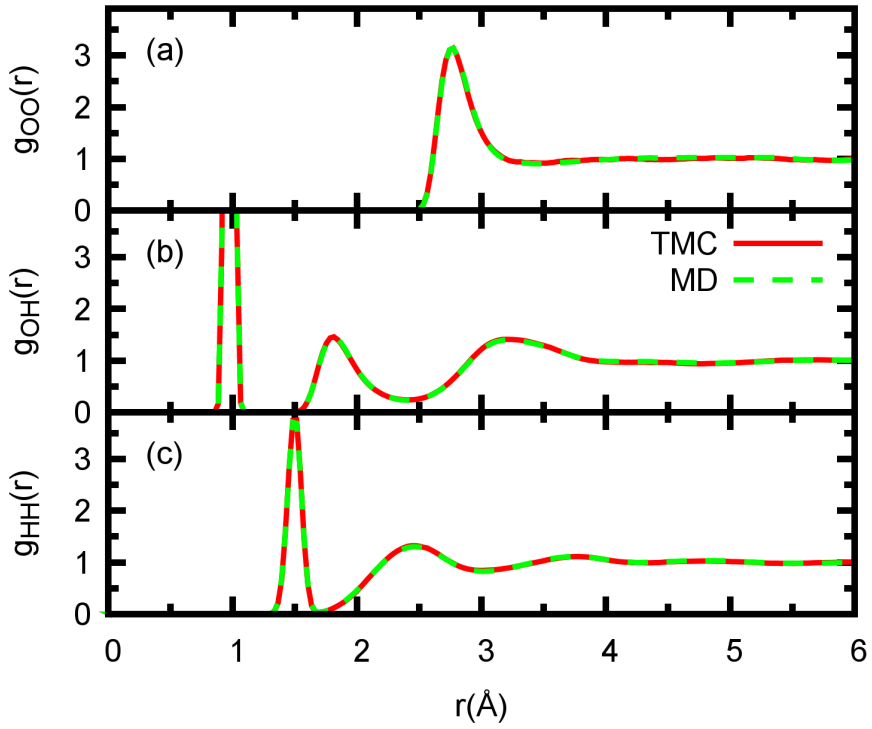


Figure 4.4: Comparison of radial distribution functions (a) O-O, (b) O-H and (c) H-H, employing 500000 steps of NVT TMC (red) and 250ps of NVT MD simulation (green). A liquid water system is investigated, consisting of 64 H_2O at 295K, employing a flexible classical potential.

4.3 Performance

In this section, the performance of the TMC algorithm (including the distinct optimizations) is compared. Therefore, the average number of Markov chain

elements sampled within a certain run time is considered or, in other words, the average calculation time for computing one Markov chain element. An efficient implementation of the TMC algorithm requires a continuous calculation of the energies (dominating the simulation) and additionally a large ratio of the number of Markov chain elements related to the number of totally computed elements. Furthermore, a low communication overhead avoids delays in the sampling process. The reached speedup ($\frac{\text{serial execution time}}{\text{parallel execution time}}$) is measured by the ratio of the reference time divided by the actual calculation time of one Markov chain element (employing average values). As reference, the average computing time of one element is employed, investigating a serial MC at comparable computing unit sizes.

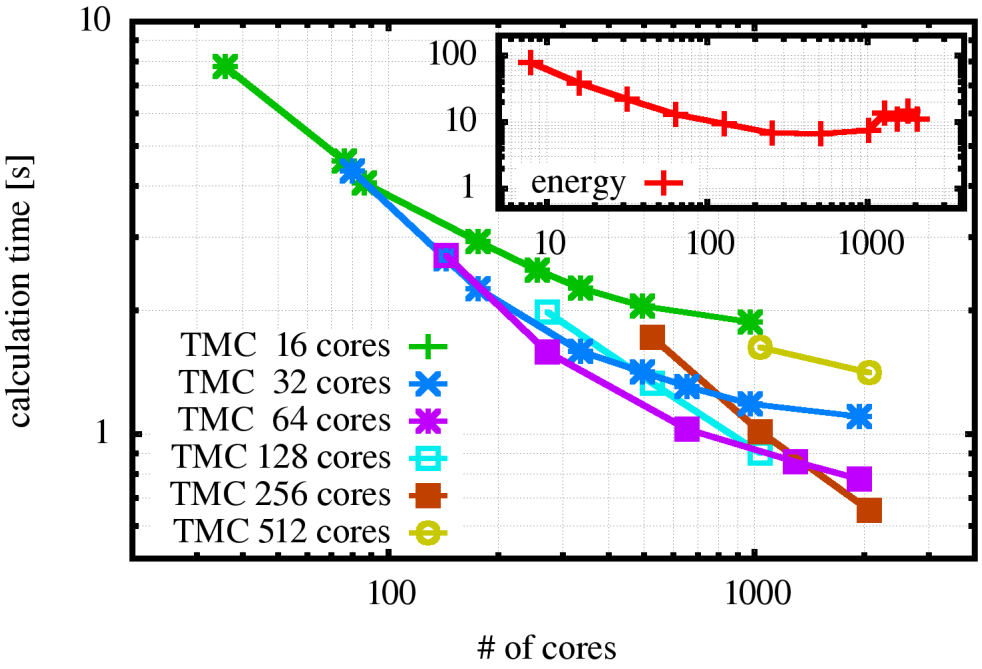


Figure 4.5: Time for a parallel energy calculation (red) compared to the average time for sampling one Markov chain element investigating group sizes (N_E) of 16 (green), 32 (blue), 64 (violet), 128 (light blue), 256 (brown) and 512 (yellow) cores. On the abscissa the total number of used cores (N_{tot}) is plotted using various numbers of energy workers

First, the optimal group size is theoretically determined by the scaling

of the underlying domain decomposition method. Therefore, the energy calculation time is investigated for an increasing size of the (single) computing unit. In the present calculations of a system (64 H_2O calculated with a PBE functional), the minimum occurs at 256 cores per group. At larger group sizes the communication overhead dominates (red curve in Fig. 4.5). Then, different TMC settings are examined, which distribute the available processors into a varying number of workers (with various group sizes). An increasing number of (energy) workers lead to similar logarithmic behaviors and increasing group sizes reflect the expected behavior (similar to the red curve in Fig. 4.5). Due to available computing resources, further calculation are investigated at group sizes of 16 cores per group, which fit in one computing node.

The implemented optimizations influence the prefactor f of the theoretical ($f = 1$) speedup of the tree algorithm

$$S = f \log(N_E) + 1. \tag{4.4}$$

utilizing N_E energy worker groups. Employing the pre-sampling method (using the approximate potential), the computing time $t_{CC} + t_E \rightarrow t_E$ (in the present case $30t_{CC} \approx t_E$) is reduced, resulting in an increased speedup ($f = 1.06$) (blue curve in Fig. 4.6). In addition, the dynamic canceling concentrates the computing power on the Markov chain, resulting in a significant reduction of the time spent on incorrectly predicted Markov chain elements, as N_E and t_E increase. Hence, the increasing number of calculated Markov chain elements within a certain time raises the prefactor of the speedup to $f = 1.39$ (blue curve in Fig. 4.6). Furthermore, the accurate prediction of the correct Markov chain direction by estimating the final energies (utilizing intermediate SCF step information) decreases the time until cancellation of wrong elements. As a result, the efficiency is increased significantly ($f = 1.77$, violet curve in Fig. 4.6). It should be mentioned that an efficient canceling procedure requires an instantaneous processing of the canceling message on the worker side. In the presented simulations the number of employed workers for configurational changes is large enough to

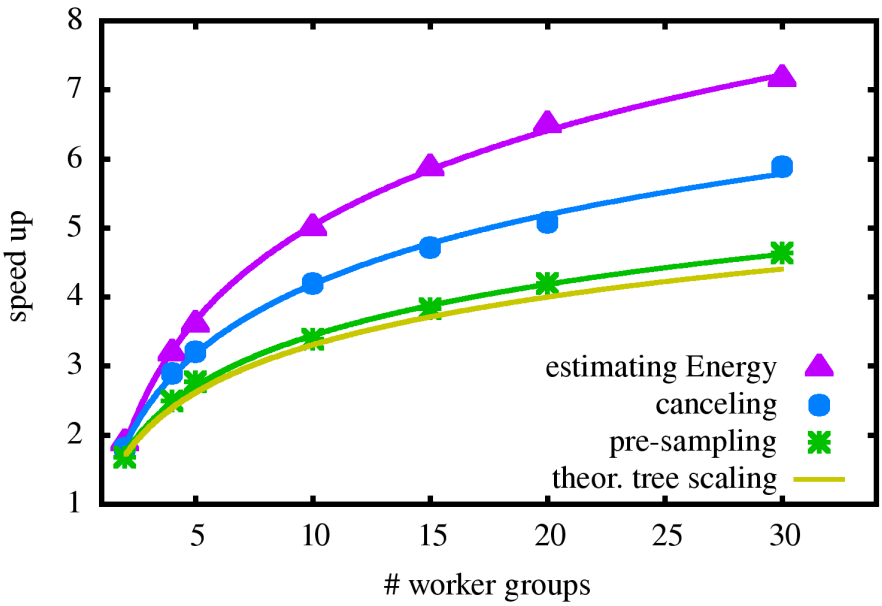


Figure 4.6: Theoretical speedup for the investigated optimizations compared to the theoretical speedup ($log(N_E)+1$) of the tree algorithm (yellow). Starting with (only) pre-sampling (green), estimations of average acceptance and canceling (blue) is added. Additionally estimations of actual acceptances (final energy estimation) (violet) are considered for investigated system of 96 H_2O using a PBE functional (with ≈ 15.7 SCF steps), sampled with the NMC method using an approximated classical potential ($30t_{CC} \approx t_E$). The reference is a single group TMC calculation without configurational preparation (comparable to serial MC).

ensure prepared configurations for every available energy worker. The selected group size of these workers is $s_{cc} = 1$ core, as the computational cost of the approximate potential is low.

The mentioned communication overhead increases with an increasing number of workers, especially on the master side, because more and more messages are passed around. The workload of the processors can be monitored by the timings for probing for new messages (`MPI_PROBE`). On one hand, the workers should be continuously loaded with the related tasks (without delays in receiving a new task), hence the time for probing should be negligible. On the other hand, the master should have enough capacity to handle

all messages instantaneously and create new tasks, which is indicated by a long time for probing. For example, a simulation on 5 groups (95 processors in total, including the master and a analyzer group), spends 335s on the master side for probing, for a total run time of 1067s. This time (spent for probing on the master side) illustrates the capacity to handle even more tasks (the number of tasks are limited by the number of workers). However, this relatively short time (335s out of 1067s) monitors the already huge number of messages, which are sent/received, especially to/from the 14 workers for configurational change. The fast creation of new configurations and relatively slow calculation of energies ($30t_{CC} \approx t_E$) increase the time spent in the tree search routines (master side). In contrast to the master, the workers spend less than 40s in the broadcast routine, mainly employed by the domain decomposition method.

To conclude, the TMC algorithm creates single Markov chains similar to a serial (regular) Monte Carlo algorithm. Due to the speculative approach of TMC, configurations can be pre-sampled, as a result the time for calculating (even computational expensive) configurational changes vanishes. Thus, a speed up of factor two can be obtained compared to the serial algorithm, for the case that the time calculating the configurational change is almost as long as calculating the exact energy. Furthermore, the basic TMC algorithm speed up the calculations logarithmically with respect to the number of worker groups compared to the serial Monte Carlo algorithm. The optimizations leads to an additional improvement of the performance of a factor of roughly two.

These scaling analysis is independent of the size of the simulated system and the method employed for calculating the energies. Both, the size as well as the method only effects the scaling of the underlying domain decomposition method, distributing each configuration within a worker group.

Chapter 5

Dielectric Properties of Water Ice, the Ice Ih/XI Phase Transition and an Assessment of DFT[39]

The dielectric properties of the hydrogen disordered hexagonal phase (Ih) of water ice have been computed using density functional theory (DFT) based Monte Carlo simulations in the isobaric-isothermal ensemble. Temperature dependent data yield a fit for the Curie-Weiss law of the system and hence a prediction of the temperature of the phase transition from the Ih phase to the hydrogen ordered ice XI phase. Direct simulations around the phase transition temperature confirm and refine the predicted phase transition temperatures and provide data for further properties, such as the linear thermal expansion coefficient. Results have been obtained with both hybrid and semi-local density functionals, which yields insight in the performance of the electronic structure method. In particular, the hybrid functional yields significantly more realistic dielectric constants than the semi-local variant, namely $\varepsilon \approx 116$ as opposed to $\varepsilon \approx 151$ at 273K ($\varepsilon_{experiment} = 95$). This can be attributed to the tendency of semi-local functionals to be biased to configurations with a large dipole moment, and their overestimation of the dipole

moments of these configurations. This is also reflected in the estimates of the Ih/XI transition temperature, which is 70-80K and 90-100K for the hybrid and semi-local functional respectively. DFT based sampling of the millions of configurations necessary for this work has been enabled by a Tree Monte Carlo algorithm, designed for massively parallel computers.

5.1 Introduction

Water ice is a ubiquitous substance and yet despite its familiarity and intensive study for centuries, some of its most basic properties are not fully understood. For example, in the 1850s, Michael Faraday proposed the surface of ice was 'liquid-like'[40], yet a detailed atomic scale understanding of this structure is still elusive. Below the onset of 'quasi-liquid layer' formation, ice at ambient temperature (ice Ih) is fully crystalline. However, although the water ice surface shows both short and long-range order, it also displays behaviour normally associated with amorphous materials [41]; surface and subsurface vacancy formation energies, which would expected to be similar, vary by ≈0.8eV, approximately three times the strength of a hydrogen bond. This observation is explicable through consideration of the effects of *orientational disorder* - where a water molecule can adopt one of six possible orientations; the orientational disorder leads to positional disorder of the hydrogen atoms on an ordered oxygen sub-lattice. This disorder contributes to the amazing richness in ice's phase diagram, giving rise to 15 known phases [42]. This hydrogen disorder is the result of a delicate balance between configurational entropy, that has been accurately estimated by Pauling[43] and Giauque[44], and the small energy differences between the various configurations[45] that respect the Bernal-Fowlers ice rules.[31] This disorder, and the associated fluctuations of the total dipole of the sample, lead to a high dielectric constant for solid water ice Ih ($\varepsilon(272K) = 95$), which is surprisingly similar to that of liquid water ($\varepsilon(278K) = 86$). The term 'dielectric constant' of ice Ih is somewhat imprecise, as the hexagonal symmetry of ice Ih implies that the dielectric tensor will have two unique eigenvalues, $\varepsilon_\perp$ and $\varepsilon_\parallel$. The former value is two-fold degenerate, with eigenvectors in the hexagonal plane and

orthogonal to the optical axis (also named c or z axis), while the eigenvector corresponding to the latter is parallel to the optical axis of the crystal. Indeed the dielectric anisotropy of ice is exploited in the characterisation of polar ice sheets by radio waves [46] as polycrystalline ice causes a distinct attenuation of radio waves in comparison to oriented ice. If all microscopic configurations of ice were energetically equivalent, as assumed in the Pauling model (see also Minagawa[47]), the anisotropy $((\varepsilon_{\parallel} - \varepsilon_{\perp})/\varepsilon_{\parallel})$ would be vanishingly small (see e.g. Aragones *et al.* [48]). Early measurements of this anisotropy, performed by Paul Scherrer and co-workers in a freezer room of a local brewery in Zurich[49], yield an anisotropy of 13% near the melting temperature. This value, later confirmed (12%) and extended to a wider temperature range by Kawada[50], suggests that the precise energetics of these configurations does matter and that the Pauling model is thus not sufficient in this context. The trigger for these experiments at low temperature was his earlier observation[51] of an indication for a phase transition near 70K to what is now known as the proton ordered, ferroelectric, hexagonal phase ice XI.[52] Such a phase transition should be clearly visible in the dielectric properties, as the ordered phase has a low dielectric constant, while the relevant component of the dielectric tensor of the high temperature phase can be approximated by a Curie-Weiss law

$$\varepsilon_{\parallel} - \varepsilon_{\infty} = A/(T - T_C). \tag{5.1}$$

where A is the Curie constant and T_C the Curie temperature, which should be similar to the phase transition temperature. These experiments are challenging and have lead to a wide range (6K-127K) of estimates for T_C, as reviewed in Ref. [32]. The origin of this difficulty is the long dielectric relaxation time in pure ice and the slow kinetics of the phase transition, which requires doping with KOH to reduce the associated time-constants from years to more managable values. The uncertain influence of the dopant and the difficulty of growing single crystals containing KOH have made characterization of this phase transition a challenge to experiment. Nevertheless, recent dielectric experiments agree on a phase transition temperature of 72K for

H_2O[53, 54] (76K for D_2O[52]), while thermally stimulated depolarization provides further evidence for ferroelectric ordering[55], and neutron diffraction has provided the crystal symmetry ($Cmc2_1$) of the ice XI phase[56, 57]. The dielectric tensor should thus be considered a quantity that is very sensitive to important aspects of water, namely the polarization of the system, and the detailed energetics of the various hydrogen bonding configurations. The fact that the structure of ice Ih, contrary to that of the liquid (see Ref. [58] for a recent discussion), is well known, also makes the system ideal to quantitatively asses the quality of simulation models. Furthermore, simulation models that reproduce the well established experimental data, for example the high temperature dielectric tensor, can than be used to investigate more controversial aspects, such as the influence of electrostatic screening between charged defects and orientational (Bjerrum) defects. It is important to recognise that force-field based approaches, including sophisticated high order multipole models cannot currently accurately capture subtle differences in energy between hydrogen orderings. Therefore, quantum mechanical based approaches are essential for benchmarking. Nevertheless, force field based approaches have provided tremendously instructive insights into ice physics. Rick and Haymet proposed a Monte Carlo move suitable for off-lattice calculations of ice[34], and demonstrated that several non-polarizable empirical atomistic models of water significantly underestimate (by 100%) the dielectric constant of ice. This result was later verified and extended by several other groups[59, 60, 48]. Adding polarizability to the model improved results significantly, which was attributed to the larger molecular dipole in these models.[34] However, none of the methods displayed a significant anisotropy of the dielectric tensor, as observed experimentally, which may related to the fact that these models predict an anti-ferroelectric ordered phase to be more stable than the experimental ferroelectric phase.[34, 60, 48] In fact, Hirsch and Ojamäe found[45] an almost anti-correlation between the energies computed with DFT and empirical methods for the sixteen unique proton ordered configurations compatible with an orthorhombic unit cell containing eight water molecules, which has been attributed to incorrect higher order electrostatic multipoles in the empirical force fields[61]. Furthermore, HF

and DFT yield (independent of the density functional employed), a lowest energy configuration that is indeed ferroelectric[45, 62, 63, 61, 64], indicating that these methods might be capable of reproducing the fine energy details in this system. Singer and co-workers parametrized an analytical model based on graphs, i.e. hydrogen bond network patterns, to DFT (BLYP) reference values, and with this approach, were able to perform Monte Carlo simulations that predict the phase transition temperature between Ih and XI to be 98K in good agreement with experiment.[62, 65] This is a very interesting and powerful approach, but is probably limited to models containing less than ≈60-70 molecules, because of the combinatorial explosion of enumerating possible hydrogen bonding arrangements. In addition, the model may not be easily extended to consider orientational (Bjerrum) or ionic defects (i.e. hydronium and hydroxide). Unfortunately, the dielectric properties were not reported, so cross validating the employed model directly is not possible. In all simulations discussed here, nuclear quantum effects, which are relatively small in this context, have not been taken into account, and hence comparison with results for D_2O, which is the more 'classical' isotope, will be considered whenever possible. Rusnak and co-workers estimated the static dielectric constants of liquid water and ice for PBE, using a free energy perturbation approach that employs a classical potential to sample configurations, which are reweighted to *ab initio* values.[66] However, obtained results underestimate the experimental value significantly ($\varepsilon(253K) = 67$). This difference might in part be due to their use of a Bader-like analysis to obtain molecular dipoles, instead of relying on the modern theory of polarization.[67, 68] The latter theory was employed in the simulations of Sharma *et al.* to compute the dielectric constant of liquid water.[69] Based on a 20ps simulation of liquid water with the PBE functional in the canonical ensemble (NVT), a value in surprisingly good agreement with experiment was obtained ($\varepsilon(330K) = 67$ vs. $\varepsilon(330K) = 68$). For ice Ih, the proton configurations could not be sampled, but a molecular dipole derived from a representative ice configuration (3.32D), in good agreement with experiment would be obtained for a plausible angular correlation factor $G_K \approx 2.55$. It is noteworthy that the same correlation factor obtained by commonly used empirical potentials varies in

the range 1.34-2.7.[48]

In the current work, the merits of the various approaches mentioned above have been combined, and explicit sampling of hydrogen disorder at the DFT level is employed to 1) compute the dielectric tensor at various temperatures using the modern theory of polarization, 2) to determine the Ih/XI phase transformation temperature directly, and 3) to assess the quality of semi-local and hybrid density functional theory for water ice.

5.2 Computational methods

Calculations have been performed with the free simulation package CP2K/ Quickstep[21, 70] that has been enhanced with a Monte Carlo (MC) algorithm, named Tree Monte Carlo (TMC). The system studied consists of 96 water molecules in an orthorhombic cell, initially $13.57\text{Å} \times 15.67\text{Å} \times 14.73\text{Å}$. This cell is sufficiently large to allow for a Gamma-point calculations, and is similar in size to the classical model systems employed previously (e.g. 128 molecules in Ref.[34]). Near the phase transition temperature, size effects can not be excluded. Monte Carlo simulations have been performed in the isobaric-isothermal (NPT) ensemble[71] at 1 bar, except when the canonical (NVT) ensemble is mentioned explicitly. DFT calculations are based on the semi-local functionals by Perdew, Burke, and Ernzerhof (PBE)[72] and by Becke, Lee, Yang and Parr (BLYP)[73, 74] and the corresponding hybrid functionals PBE0[75] and B3LYP[76, 74, 77]. The latter two functionals contain a fraction of Hartree-Fock exchange (denoted HFX), 25% and 20% respectively, and are thus non-local. All functionals are combined with a dispersion correction by Grimme (which we denote D2) [78], which is essential to obtain a qualitatively correct density of liquid water in NPT simulations.[79] Norm-conserving Goedecker-Teter-Hutter (GTH)[80] pseudopotentials are employed together with a triple-zeta valence basis set and two polarization functions (TZV2P). HFX is computed with a robust gamma-point implementation[81, 82] using a truncation radius of 6Å, and the Auxiliary Density Matrix Method (ADMM) with the pFIT3 basis.[83] Periodic boundary conditions have been used throughout, and electrostatics

are computed using an Ewald-sum. This implies so called conducting or tin-foil boundary conditions[84], suitable for samples embedded in a medium of high dielectric constant, and is equivalent to enforcing closed circuit electrical boundary conditions without applied bias.[85] These boundary conditions are almost always applied in electronic structure calculations and typically used in force field based simulation of ice (see e.g. [34, 48, 60]). The plane wave cutoff was 800 Ry for NPT and 400 Ry for NVT calculations, with the number of grid points kept fixed during NPT simulations.[70, 86]

The elements of the dielectric tensor are obtained as the second moments of the dipole distribution using

$$\varepsilon_{\alpha\beta} = \varepsilon_\infty + \left(\frac{4\pi}{3Vk_BT}\right)\left(\langle \mathrm{M}_\alpha \cdot \mathrm{M}_\beta\rangle - \langle \mathrm{M}_\alpha\rangle\langle \mathrm{M}_\beta\rangle\right), \tag{5.2}$$

where M_α is the component of the total cell dipole in the α direction, and ε_∞ is the optical dielectric constant, taken to be one in the following, and $\langle\cdot\rangle$ denotes ensemble averaging. In the context of periodic DFT, changes in the cell dipole are obtained from the Berry phase formulation of polarization[67].

Monte Carlo generates a sequence of configurations, a Markov chain, that samples a given ensemble based on acceptance/rejection of random configurational changes (moves). Three important strengths of Monte Carlo are exploited in the context of this study. First, in addition to the traditional moves (atom and molecule translation, molecule rotation, and non-isotropic volume changes), a proton reordering move is employed[34]. This specialized move allows for re-orienting several molecules in a closed loop of hydrogen bonded neighbors in a way that obeys the Bernal-Fowler ice rules and may change the effective dipole of the unit cell, circumventing the high energy barriers that the physical process must overcome in experiment or unbiased molecular dynamics simulations. Second, in the context of DFT it is advantageous to employ a pre-sampling strategy in which a sub-Markov chain is generated with an approximate classical potential.[24, 20, 86] In this nested MC approach, the entire sub-Markov chain is accepted/rejected with a single DFT calculation, leading to exact sampling of the DFT potential at much reduced computational cost. Note that converged MC results are independent

of the approximate classical potential, but the rate of convergence depends on the quality of the approximation. Here a refitted classical non-polarizable potential based on the model from Ref.[87] has been employed, which provides good acceptance. In our current approach, 50 (NPT) or 100 (NVT) nested MC moves are applied, which are constrained to a fixed sub-cell with edges of length 8Å in order to retain sufficiently large acceptance rates in the DFT step. This nested MC setup leads to a high fraction of the DFT calculations being employed to accept/reject hydrogen reordering moves, resulting in effective sampling of the cell dipole. Third, Monte Carlo contains an intrinsic parallelism that has so far not yet been exploited in molecular simulation, but is here used by the TMC algorithm for obtaining long Markov chains on massively parallel computers. It is based on the observation that new (future) configurations in the Markov chain can be constructed on the fly and instantaneously, before the energy of the current configuration has been evaluated, as neither energy nor forces are required for the MC moves.[17] This approach requires speculation on the outcome of the acceptance check, or when used in a systematic fashion, builds a tree of configurations assuming both possible outcomes of this check. All the configurations present in the tree can be computed simultaneously, limited only by the available resources. Finally, the Markov chain, identical to the conventional serial one, is constructed, discarding the configurations that are off the chain, as the energies become available. The implementation of this algorithm, which includes a number of techniques to reduce the amount of discarded work, is freely available[21] and will be discussed in more detail elsewhere. TMC has been key to reduce the wall time per generated Markov chain element to 2.5 s and 14 s for PBE and PBE0 respectively, and ultimately allowed for single MC chains with nearly a million moves.

5.3 Results and discussion

The temperature dependent dielectric constant of ice Ih at ambient pressure is a central result of this work. It is a challenging quantity to obtain by simulation, as extensive sampling is necessary for convergence. However, provided

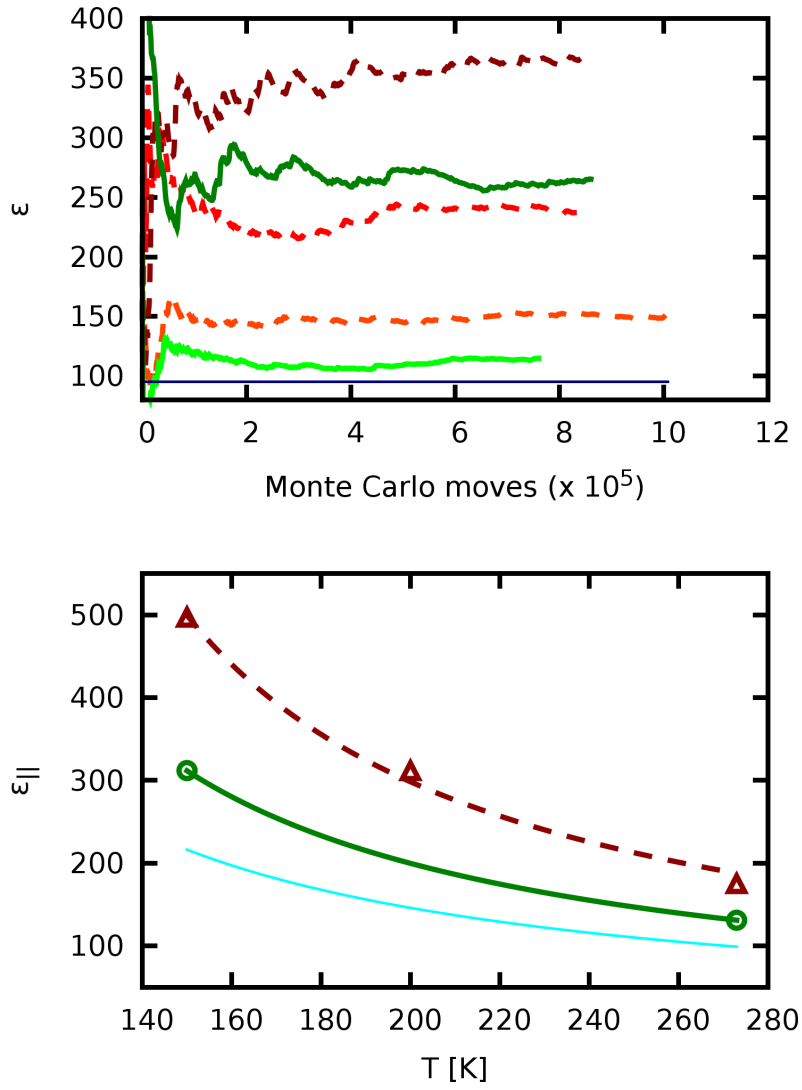


Figure 5.1: upper panel: Shown are the running averages of ε for PBE-D2 at 273K (orange), 200K (red), 150K (brown) and PBE0-D2 at 273K (lightgreen) and 150K (green). The experimental value of Johari [32] at 273K is shown in blue. lower panel: Shown are the converged values of $\varepsilon_\perp$ for PBE0-D2 (circles) and PBE-D2 (triangles), with the corresponding fit (solid lines) of Curie-Weiss law (5.1). Experimental data of Kawada *et al.* [53] (light blue) are shown for comparison.

100000s of MC steps are performed, convergence can be reached even with

Table 5.1: Elements of the dielectric tensor and dielectric anisotropy obtained with PBE-D2 and PBE0-D2 in NPT ensemble. The difference between ε_{xx}, ε_{yy} and $\varepsilon_\perp = (\varepsilon_{xx}+\varepsilon_{yy})/2$ is used as an estimate of the statistical uncertainty.

	T[K]	10^5 MC moves	ε	$\varepsilon_\perp$	$\varepsilon_\parallel$	$(\varepsilon_\parallel - \varepsilon_\perp)/\varepsilon_\parallel$
PBE0-D2	273	7.6	116	108 ± 3	131	17%
PBE0-D2	150	8.6	266	234 ± 0	331	30%
PBE-D2	273	10.1	151	139 ± 7	174	20%
PBE-D2	200	8.3	238	202 ± 7	310	35%
PBE-D2	150	8.4	366	298 ± 28	502	41%

first principles simulations. This is demonstrated for five MC simulations in the NPT ensemble, based on either the PBE-D2 or PBE0-D2 functional, in the upper panel of Fig. 5.1. The relevant components of the dielectric tensor obtained by these simulations are summarized in Table 5.1. This data clearly shows that ε obtained with PBE-D2 is significantly larger than the one obtained with PBE0-D2. At the melting point, $\varepsilon(273K) = 116$ is in fair agreement with the experimental result of Johari *et al.*[32] $\varepsilon(273K) = 95$, while PBE-D2 significantly overestimates it ($\varepsilon(273K) = 151$). The origin of this difference is related to the nature of the density functionals, as we discuss in more detail below. Of particular interest is the value of the dielectric anisotropy ($(\varepsilon_\parallel - \varepsilon_\perp)/\varepsilon_\parallel$), because it is indicative of the XI/Ih phase transition. Simulations based on empirical force fields predict very small or zero values of the anisotropy.[60, 34, 48] PBE0-D2 calculations at 273K result in an anisotropy value of 18%, which is slightly greater than the measured value of approximately 12% of Kawada *et al.*[50]. PBE-D2 overestimates the anisotropy more significantly (22%), indicative of a bias towards configurations that display a large dipole along the $\parallel$ direction. Note that, the average *molecular* dipole moments found for apolar ice slabs computed with PBE were approximately 10% greater than PBE0[41] which contributes to the observed difference in anisotropy using the semi-local and hybrid functionls. As the temperature decreases and the Ih/XI phase transition temperature approaches, the anisotropy becomes larger. A fit of the Curie-Weiss law through the simulation data, shown in the lower panel of Fig. 5.1 yields $T_C = 60$K and $T_C = 79$K for PBE0-D2 and PBE-D2, respectively. However,

given the limited data, the uncertainty in this fit is likely large, on the order of 10K. Furthermore, a true divergence as described by equation Eq. 5.1 can not be observed due to finite size effects. Nevertheless, the Curie temperatures found are close to the experimental phase transition temperature; 72K for H_2O and 76K for D_2O.

Based on these estimates, the phase transition temperature has been determined by direct simulation. In particular, NPT MC runs at various temperatures between 50 and 100K have been started from the same ice Ih configuration, and run for at least 250000 MC steps. As shown in Fig. 5.2, this leads to a spontaneous phase transformation to the expected ice XI phase for the low temperatures, while the systems at the higher temperatures remain in the ice Ih phase. The resulting phase transition temperatures can thus be bracketed by 70-80K (PBE0-D2) and 90-100K (PBE-D2), in fair agreement with the estimates derived from the Curie-Weiss fit. The PBE-D2 result is in agreement with the 98K computed by a graph based model parametrized to the semi-local BLYP functional, yielding additional support for this approach.[62, 65] Nevertheless, the PBE0-D2 model can be assumed to be the most predictive, given its favorable comparison to dielectric experimental data, obtained with well established experiments near the melting point. Furthermore, the fact that the transition temperature is relatively insensitive to the precise DFT model suggests that this prediction will be robust.

The advantage of direct MC simulation of a reliable atomistic model is that further properties can be obtained from the generated trajectories. Here we investigate the temperature dependent density, or equivalently the volume, given its relevance for ε (see below) and the fact that this quantity is directly accessible from the NPT simulations performed. Indeed, NPT simulations employ a flexible simulations cell, which fluctuates around a temperature dependent average value ($\sigma(273K) \approx 0.12$Å for our simulation cell). At the melting point, both models overestimate the density, with 0.960 and 0.995 g/cm^3 for PBE0-D2 and PBE-D2 respectively, compared to the experimental results obtained by Röttger[88] (0.917 g/cm^3). PBE0-D2 is again the best model, with a volume that is underestimated by 4.5 %. The tem-

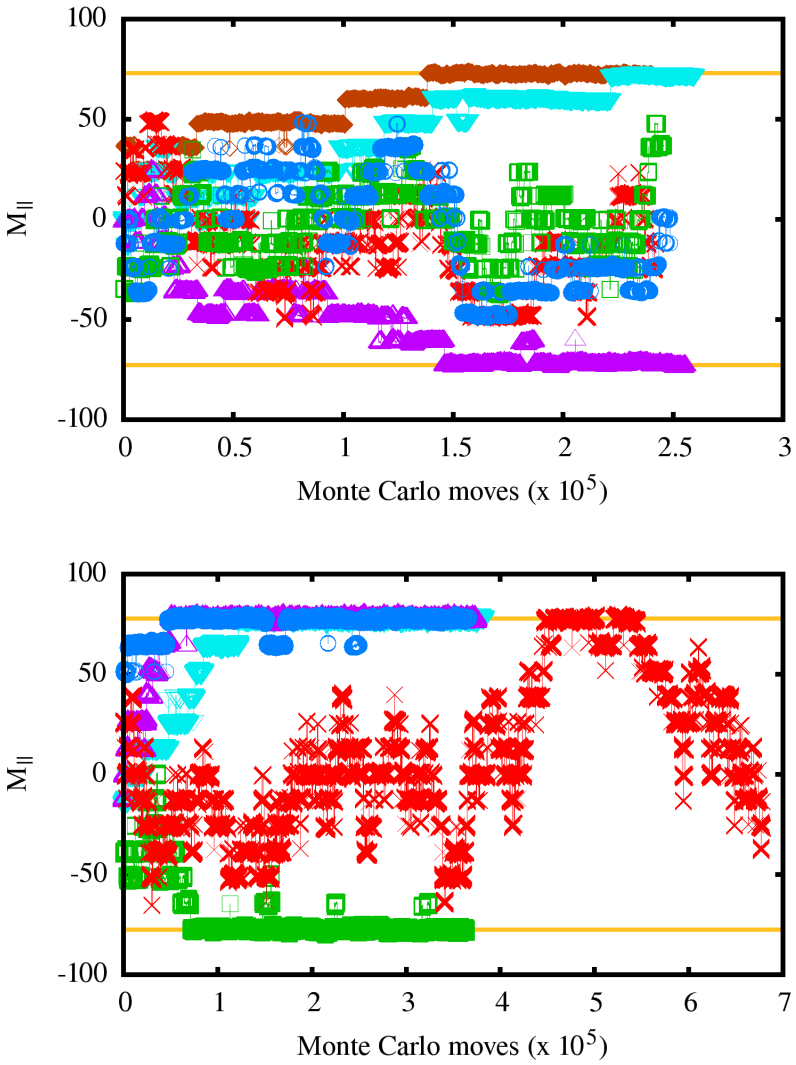


Figure 5.2: Evolution of the z component of the total cell dipole for NPT calculations at various temperatures. upper panel: PBE0-D2, lower panel: PBE-D2. The different colors and symbols correspond to 100K (red, crosses), 90K (green, squares), 80K (blue, circles), 70K (violent, triangles up) and 60K (light blue, triangles down) and 50K (brown, diamonds). The solid (orange) lines indicate the permanent dipole of the ice XI phase.

perature dependence of the volume can be described by the linear thermal

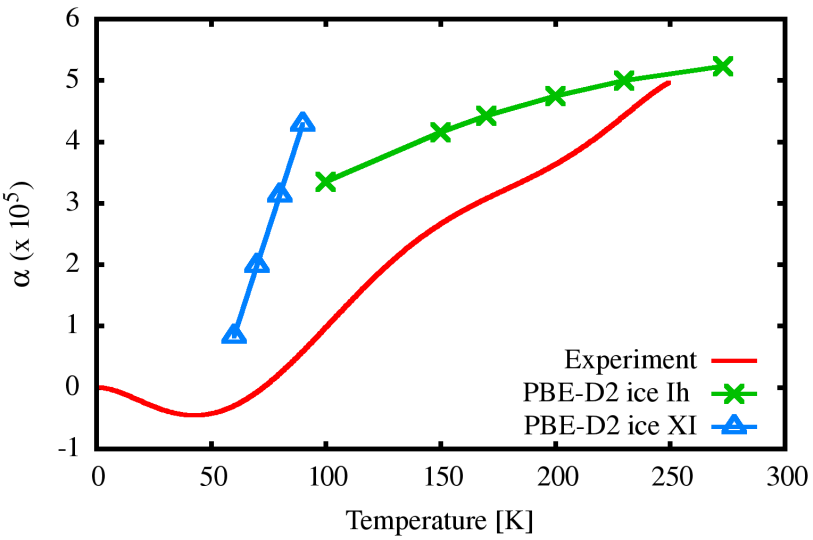


Figure 5.3: Linear expansion coefficient α, fitted using NPT ensembles calculated with and PBE-D2 functionals for ice Ih phase (crosses) and ice XI phase (triangles), compared with experimental results of D_2O of Röttger[88] (solid red line)

expansion coefficient $\alpha(T)$ defined by

$$\alpha(T) = \frac{1}{3}\frac{1}{V(T)}\frac{\partial V(T)}{\partial T}. \tag{5.3}$$

Only for PBE-D2 are sufficient data available to yield a stable polynomial fit of the temperature dependence of the volume, using a third order polynomial for the ice Ih phase (100-273K) and a second order polynomial for the ice XI phase (60-90K). In Fig 5.3, the corresponding $\alpha(T)$ is compared to the experimental results by Röttger[88]. Given the slow kinetics of the XI/Ih phase transition, it is reasonable to assume that the experimental system consists of a mixture of both phases[89], with a precise constitution that depends on the preparation method. Indeed, ice XI has never ever been reported to have been made more than 70% pure.[52] It is therefore reasonable to assume that the experimental curve can be approximated by a combination of the simulation data of the pure phases, which suggests good qualitative

agreement. In particular, for even lower temperatures, a negative thermal expansion coefficient could be expected.

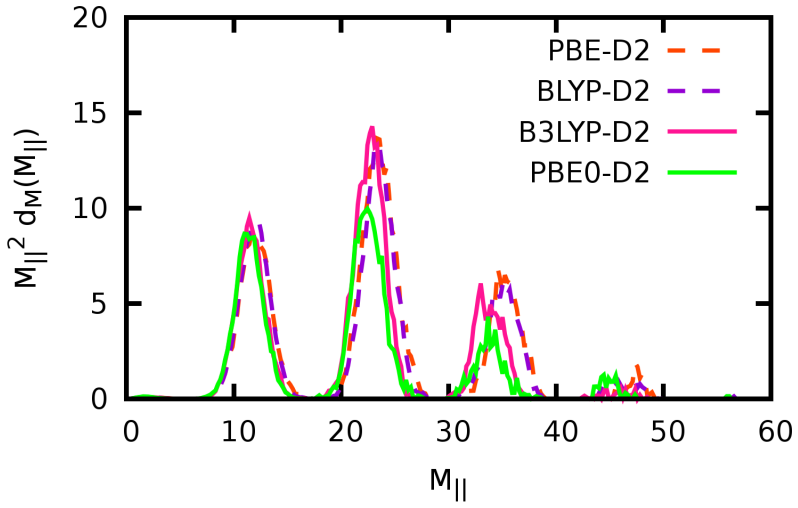


Figure 5.4: Distribution of the z component of the dipole moment, weighted by the dipole squared. This weighting corresponds to the contribution of these peaks to ε. Results have been computed for semi-local (PBE-D2 and BLYP-D2) and hybrid (B3LYP-D2 and PBE0-D2) functionals in the NVT ensemble at 273K.

Table 5.2: Elements of the dielectric tensor and the dielectric anisotropy obtained in the NVT ensemble at 273K and experimental density with the PBE-D2, PBE0-D2, BLYP-D2 and B3LYP-D2 functionals. To illustrate the effect of compression, one additional NVT result at 106% of the density is provided for PBE-D2.

	T[K]	10^5 MC moves	ε	$\varepsilon_\perp \pm \frac{(\varepsilon_{xx}-\varepsilon_{yy})}{2}$	$\varepsilon_\parallel$	$(\varepsilon_\parallel - \varepsilon_\perp)/\varepsilon_\parallel$
Exp. (Ref. [32])	273		95			
Exp. (Ref. [50])	273		90	87	99	12%
PBE-D2	273	6.9	127	114 ± 2	152	25%
PBE-D2 (106%)	273	3.2	145	126 ± 3	183	31%
PBE0-D2	273	2.8	103	98 ± 4	112	12%
BLYP-D2	273	13.7	122	110 ± 1	144	24%
B3LYP-D2	273	3.2	110	98 ± 1	133	26%

Finally, we show that the observed differences between PBE-D2 and

PBE0-D2 are more general in nature. To do so, we have performed MC simulations at constant volume (NVT), 273K and experimental density, and compared two pure DFT (PBE-D2 and BLYP-D2) and two hybrid (B3LYP-D2 and PBE0-D2) functionals. The data for these simulations is summarized in Table 5.2. Constant volume simulations facilitate the comparison between the functionals, and can be used to highlight the importance of employing the NPT ensemble for the calculation of the dielectric constant. In particular, the PBE-D2 simulations have been performed at 100% and 106% of the experimental density, the latter being similar to the equilibrium density of the PBE-D2 functional. While the latter simulation yields $\varepsilon = 145$, very similar to the NPT results of $\varepsilon = 151$, the simulation at experimental density yields $\varepsilon = 127$, a significant reduction of the value. Compared to the NPT simulations, a similar reduction is also observed for PBE0-D2, which at experimental density provides an even better estimate of ε (103) and anisotropy (12%). This sensitivity of ε to the density (volume), suggests that modifications to the functional that improve the lattice parameters could yield enhanced dielectric properties. At the same density, the overestimation of ε in the case of semi-local functionals is due to two effects, visible in Fig. 5.4. These effects are more important than the small differences in geometry between semi-local and hybrid functionals (e.g. $r_{OH} = 1.000$ and $r_{OH} = 0.985$ respectively). First, as can be seen as a shift of the position of the peaks, dipoles of a given configuration are enhanced by roughly 5%. Second, visible by the area under the peaks, these polarized configurations are sampled more often, i.e. are energetically favorable. Both effects are consistent with the overestimation of polarizability by semi-local density functionals (see also Ref.[41]) and with the improvements expected for a hybrid functional.[90] Ultimately, this is a result of the underestimation of the band gap by semi-local functionals, a deficiency that is, also in water ice, significantly improved by hybrid functionals.[64]

5.4 Conclusion

The temperature dependent dielectric constant of ice Ih contains a wealth of information about the interactions between water molecules. It contains the most visible signature of the Ih/XI phase transition, which is the result of a delicate balance between energy and entropy. Here, this property has been computed directly from extensive sampling of hybrid and semi-local DFT models in the appropriate ensemble. We find that semi-local DFT significantly overestimates the dielectric constant ($\epsilon(273K) \approx 150$ for PBE-D2), whereas hybrid density functionals ($\epsilon(273K) \approx 116$ for PBE0-D2) provide better agreement with experiment ($\epsilon(273K) \approx 95$). The electronic overpolarization and the energetic bias towards configurations with large dipole both contribute to the error observed for the semi-local functional. This effect is also observed for a different family of functionals (BLYP-D2 and B3LYP-D2), and is thus likely of general nature. Furthermore, we expect that this error will also be present in the liquid phase, which could have implications for example for simulations of the behavior of water near (electrified) interfaces, or for the solvation and interaction of ions including incipient species such as the Zundel ion and processes involving proton transport. The Ih/XI phase transition temperature has been determined both from a Curie-Weiss fit and from direct MC simulations for both PBE0-D2 and PBE-D2. All these values are in good agreement, lending support to our methodology, and yield a transition temperature of 70-80K (PBE0-D2) or 90-100K (PBE-D2). Given the validation of our model via the dielectric constant, we expect the PBE0-D2 model to be most predictive. This PBE0-D2 result confirms accurately the measurements by Kawada (Ref. [53]) and thus suggests that the experimental procedure of KOH doping can indeed be used to accelerate the kinetics of the Ih/XI phase transformation without influencing the final equilibrium structure formed. Finally, MC sampling of explicit atomistic models yields data that can be used to compute several properties, and as an example, the linear thermal expansive coefficients of ice Ih and XI has been reported in their respective temperature ranges. This shows that the steady increase in computer power and the development of novel algorithms have made the ex-

tensive sampling of disordered solids at the DFT level possible, and that this approach can now be used to evaluate properties that have long challenged experiment and theory alike.

Chapter 6

Bulk liquid water at ambient temperature and pressure from MP2 theory[91]

MP2 provides a good description of hydrogen bonding in water clusters and includes long range dispersion interactions without the need to introduce empirical elements in the description of the interatomic potential. To assess its performance for bulk liquid water under ambient conditions, an isobaric-isothermal (NpT) Monte Carlo simulation at the second-order Møller-Plesset perturbation theory level (MP2) has been performed. The obtained value of the water density is excellent (1.02 g/mL) and the calculated radial distribution functions are in fair agreement with experimental data. The MP2 results are compared to a few density functional approximations, including semi-local functionals, hybrid functionals, and functionals including empirical dispersion corrections. These results demonstrate the feasibility of directly sampling the potential energy surface of condensed phase systems using correlated wavefunction theory, and their quality paves the way for further applications.

6.1 Introduction

Understanding the structural and electronic properties of liquid water at ambient conditions is a major challenge in condensed matter simulations. Water is a crucial ingredient for a large variety of systems of prime importance in basic chemistry, biology, and physics, as well as in the applied fields of catalysis and energy production. The water molecule has a large dipole moment and polarizability, is a multiple hydrogen donor and acceptor and can easily build network structures. The total cohesive energy in the condensed phase is, as a consequence of these properties, a sum of many weak interactions. Theoretical models face therefore the challenge to describe many different effects and their subtle interplay at a high precision. The development of sophisticated empirical potentials for water [92, 93, 94, 95, 96, 97, 98, 99, 100, 101], allowed to gain insights into water's behavior and its thermodynamic properties [102, 103, 104], such as, density maxima, heat capacity and effects of supercooling. However, empirical models lack transferability and might fail if used under conditions away from their fitting range. Most importantly, as soon as water takes an active role in a chemical process, either as a strongly interacting solvent, or for example as a source of protons, the electronic properties of the water molecule need to be taken into account. In this respect, first-principles methods offer the possibility to describe all the underling physics on the same footing, simplifying the treatment of intra- and intermolecular interactions. The capability to reproduce properties of complex systems such as liquid water can therefore be used to judge the sophistication and predictive power of a given quantum mechanical model. Density functional theory (DFT) is the most used quantum mechanical method employed for studying physical and chemical properties of condensed phase systems. Many DFT based simulation of bulk water have been reported in the literature, and in this context three main methods of sampling the phase space can be recognized [105]: the Car-Parrinello molecular dynamics (CPMD) and related variants [106, 107, 108, 109, 110, 111, 112], Born-Oppenheimer molecular dynamics (BOMD) [113, 114, 115, 116, 79] and Monte Carlo (MC) sampling. [86]

Most of the CPMD and BOMD simulations were carried out in either the microcanonical (NVE) or canonical (NVT) ensemble by constraining the volume to reproduce the experimental density ρ. First principles MD simulations in the isobaric-isothermal (NpT) ensemble are much less common [79, 117], in part due to the need for different integration schemes, the increased requirements on plane waves basis sets in variable cell simulations, and the long simulation time that is needed to equilibrate and to sample volume fluctuations. Moreover, in the MD sampling of the NpT ensemble, the calculation of the virial and thus stresses is required to change the volume as a response to the imbalance between internal stress and external pressure. [118] In this respect, the appealing feature of the MC method is that thermodynamic constraints are explicit included into the acceptance rule for each trial move, solely based on the energy. This allows to perform simulations in different ensembles, *e.g.* NpT, in a relatively straightforward manner. On the other hand, an efficient sampling of phase space in MC requires smart and system dependent trial moves [119, 120, 121, 19, 122], making the application of the method more intricate than molecular dynamics where configurational sampling follows a general principle. Within the framework of MC, McGrath and coworkers [86] reported the first results from first-principles simulations of liquid water in the isobaric-isothermal ensemble at ambient pressure. These NpT-MC simulations performed at ambient conditions using the Becke-Lee-Yang-Parr (BLYP) [123, 124] functional gave a significantly less dense (about 20%) liquid than observed experimentally. These results were confirmed and extended by Schmidt *et al.* [79] using very similar computational setups, but within the framework of NpT-MD. In the latter work, in addition to BLYP, also the Perdew-Burke-Ernzerhof (PBE) [125] functional has been tested, and, in both cases, the influence of an empirical dispersion correction (DFT-D) [126] has been investigated. It was shown that pure PBE and BLYP are indeed underestimating the water density, but the inclusion of the dispersion correction significantly improves the results giving a density as well as oxygen-oxygen radial distribution function (RDF) closer to the experimental data. From these calculations the important role played by the van der Waals interactions in bulk water clearly emerged, and it was shown

that this missing interaction in standard local functionals can be efficiently included using empirical corrections. [126, 127, 128] Furthermore, it became clear that NpT simulations are essential to quantify the quality of the intermolecular interaction potential and that a fixed simulation volume should be considered an influential constraint. Despite this progress, a truly first principles simulation of liquid water in the NpT ensemble is still missing.

Here, we present the results of MP2 based NpT-MC simulation of liquid water at ambient conditions. The second-order Møller-Plesset energy [129, 130] is an effective correction to the Hartree–Fock (HF) ground state energy that accounts for electron correlation effects. It recovers a relatively large part of the dynamic correlation, while the HF ground state ensures the inclusion of exact non-local exchange. Most notably, MP2 introduces dispersion interactions in a completely non-empirical way. For water clusters, MP2 produces results that are in excellent agreement with coupled cluster theory [131, 132, 133, 134], and is one of the best methods for the WATER27 benchmark. [135] However, the advantages of MP2 come at a computational cost that is high compared to traditional DFT using local functionals. Furthermore, MP2 calculations need larger basis sets than DFT to reach a similar state of convergence. Therefore, only few applications of MP2 to condensed phase systems have been reported (see e.g. [136, 137, 138]) and condensed phase sampling has not yet been performed. The extensive calculations reported here have become possible by combining large computer resources with an efficient algorithm for the MP2 energy calculation, which we named the Resolution of Identity Gaussian and Plane Wave (RI-GPW) approach [139, 140] and implemented in the CP2K [21] program. In addition to the new MP2 results, we report also results of NpT-MC simulations obtained from a selection of density functional approximations, namely, BLYP, PBE and PBE0 [141], including empirical dispersion corrections of the D3 type [128], which have been generated to validate the approach.

6.2 Computational Details

All calculations presented have been performed with the CP2K program. [21] Hartree–Fock and DFT energy calculations employ the Gaussian and Plane Wave (GPW) method [142, 143] that makes use of a Gaussian basis to expand molecular orbitals and an auxiliary plane wave basis for the expansion of the electronic density. This dual representation allows for evaluating the Hartree contribution to the Kohn–Sham (KS) matrix in linear scaling time at full accuracy. [143] In order to have an efficient expansion of the density in plane waves, core electrons are replaced by pseudopotentials. We use dual-space pseudopotentials of the Goedecker-Teter-Hutter (GTH) type [144] specifically parameterized for the various methods used (HF, BLYP, PBE, PBE0). The non-local HF exchange calculations have been performed employing a robust Γ-point implementation. [145, 82] To avoid divergences in the non-local HF exchange energy, the Coulomb operator is truncated at half the cell length. This allows for stable calculations in periodic boundary conditions without loss of accuracy if a truncation radius larger than $5 - 6\,\text{Å}$ is used for wide bandgap systems. [82, 146] In the case of the hybrid functional (PBE0), the calculations have been performed using the auxiliary density matrix method (ADMM) [147] that allows for an approximate calculation of the non-local exchange energy at much reduced cost. This is achieved by introducing a small auxiliary Gaussian basis that is used to expand the density matrix employed in HF exchange energy calculation. The exchange energy is then corrected for the difference between the exact and approximated density using a local exchange functional. The calculation of the MP2 energy within the RI-GPW [140] approach (RI-MP2) is closely related to the original GPW method. Here, the dual representation of the electronic density is applied to the fitting density arising from the resolution of identity approximation [148, 149, 150], where the RI fitting densities are obtained by introducing an auxiliary Gaussian basis and employing the Coulomb metric. [151]

The model system consists of 64 water molecules in a cubic simulation cell under periodic boundary conditions (PBC). Initial configurations were

taken from a system extensively equilibrated using the BLYP-D3 functional at $T = 295$K and experimental density $\rho = 1.0\,\mathrm{g/cm^3}$, i.e. constant volume with cell edges of $L = 12.42\,\text{Å}$. All further simulations have been obtained with thermodynamic constraints set to ambient conditions, that is, $T = 295$K and $p = 1$bar. The employed basis set for all calculations is of the cc-TZVP form. [139, 140] That is a valence triple-zeta correlation-consistent type basis [152, 153], but specifically generated for pseudopotentials. The cc-TZVP basis consists of sets of $(5s5p2d1f)$ and $(5s2p1d)$ primitives contracted to $(3s3p2d1f)$ and $(3s2p1d)$ functions for oxygen and hydrogen, respectively. The inclusion of the f and d functions is necessary in order to obtain accurate results at the MP2 level. At the DFT/HF level this basis is of similar quality as the QZV3P basis, that has been shown to provide very well converged DFT results. [115, 154] The associated RI auxiliary basis (RI-cc-TZVP) [140] has been generated following the procedure proposed by Weigend *et al.* [150]. It is about 2.5 times larger than the primary basis and contains up to g-type and f-type functions for O and H, respectively. In the PBE0 case an additional auxiliary basis is needed for calculating the HF exchange energy with ADMM. This basis (pFIT3) contains only non-contracted Gaussian functions with primitive patterns $(3s3p1d)$ for O and $(3s1p)$ for H. The number of grid points used in the GPW scheme for the representation of the electronic density is kept constant during volume changes. The original grids are constructed for the reference cell with density $1.0\,\mathrm{g/cm^3}$ ($L = 12.42\,\text{Å}$). The charge-density cutoffs (E_{cut}) for the HF and DFT calculations are 500 and 800 Ry, respectively, while for the RI-MP2 calculation a cutoff of 300 Ry is used. The higher cutoff for DFT calculations is necessary as the exchange-correlation functional is integrated on the same grid. [143, 139] We have verified for the BLYP functional that increasing E_{cut} from 800 Ry to 1200 Ry does not affect final results. With this setup, the model consists of 64 molecules (192 atoms), 256 electrons in 128 occupied orbitals, 57 primary basis functions per molecule (3648 basis functions in total), 136 auxiliary basis functions per molecule (8704 auxiliary basis functions in total), and plane wave grids of sizes 216^3, 180^3, and 135^3.

For all the theoretical models presented (DFT or RI-MP2) the MC set-

tings are kept the same. The considered trial moves are: (1) atom translations, (2) molecular translations, (3) molecular rotations and (4) volume changes. The MC efficiency is improved with the presampling of moves. [121, 119] In this method, an additional inexpensive (approximated) potential is introduced and used to generate a short sequence of MC moves. The entire sequence is then accepted or rejected completely, based on the difference between the exact and approximated potential functions. In the present study, the presampling consists of 30 steps with 8 simultaneous moves of type (1), (2), or (3), selected randomly with equal probability. Volume moves are not considered for the presampling, but instead directly based on the exact potential. The ratio of volume move to presampling sequences is set to 1:9. After extensive testing, the maximum displacements for moves (1) to (3) was adjusted to give acceptance probabilities between $15-20\%$ for the inner moves of the presampling sequence. This choice allows for large enough configuration changes while still keeping a good acceptance rate for the exact potential. On average, 38% of the atoms in the unit cell are moved by this presampling sequence. The maximum displacement of the volume moves is tuned in order to give a 50% acceptance rate. For the DFT simulations, a classical but refitted force field based on Ref. [155] is used as approximate potential. For the MP2 simulations, a higher quality presampling based on a semi-local density functional method is employed. In particular, the PBE1W [156] functional is used with a double-zeta plus polarization basis set and including the D3 [128] empirical dispersion correction. Finally, the length of each MC simulation is reported in terms of Monte Carlo cycles. Each cycle refers to a single exact energy calculation, which thus involves either a volume move or an entire sequence of presampling moves, that can be accepted or rejected.

6.3 Discussion

A central result of this work is the computed value of the density of liquid water. This quantity is obtained from averaging the fluctuating instantaneous density as obtained during the MC simulation. The instantaneous density

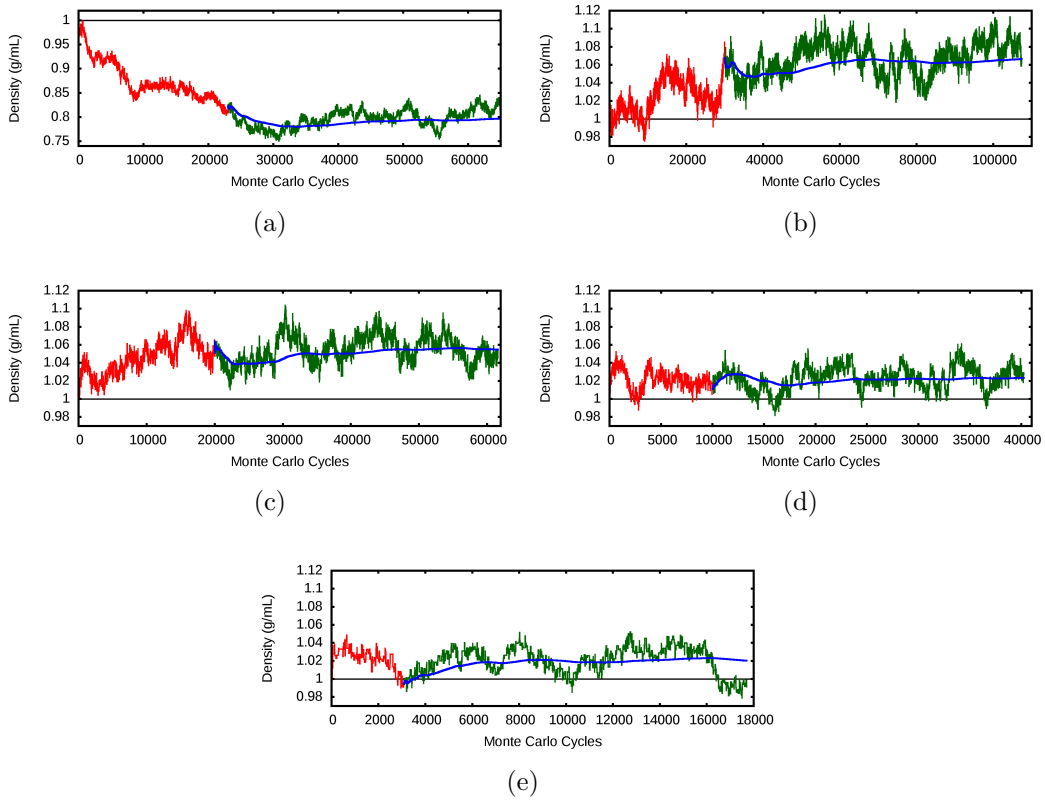


Figure 6.1: Fluctuation of the instantaneous density as a function of the Monte Carlo cycles for the NpT simulation at the various level of theory considered ($T = 295$K and $p = 1$bar). The red portion of the plots denotes the cycles considered for equilibration, the green parts refers to that used for the calculation of the average properties, the blue line shows the running average density. In each plot the horizontal black line represents the experimental density. Note the different ordinate scale in panel (a). (a) BLYP, (b) BLYP-D3, (c) PBE-D3, (d) PBE0-ADMM-D3, (e) RI-MP2.

and the corresponding running average are depicted in 6.1, while the calculated average value with the associated root mean square deviations (RMSD) are reported in 6.1. Our MP2 result for the density of liquid water at ambient conditions is ≈ 1.020 g/mL. We consider this to be in excellent agreement to the experimental value, since it implies an error in the lattice parameters of less than 1%. We attribute this to the known quality of MP2 for describing

	Density [g/mL]		1^{st} Max		1^{st} Min		2^{nd} Max		CN
	ρ	RMSD	r [Å]	$g_{OO}(r)$	r [Å]	$g_{OO}(r)$	r [Å]	$g_{OO}(r)$	n_{OO}
a	0.797	0.018	2.83	2.44	3.46	0.35	4.65	1.10	3.1
b	1.066	0.018	2.78	3.01	3.51	1.00	4.37	1.18	6.2
c	1.055	0.015	2.73	3.24	3.15	0.73	4.43	1.28	4.1
d	1.023	0.013	2.74	3.23	3.30	0.67	4.44	1.27	4.5
e	1.020	0.015	2.76	3.12	3.32	0.73	4.41	1.23	4.6
f	1.00	-	2.80	2.57	3.45	0.84	4.5	1.12	4.3

Table 6.1: Average density and structural data obtained from the MC NpT simulations at the various level of theory (BLYP (a), BLYP-D3 (b), PBE-D3 (c), PBE0-ADMM-D3 (d), RI-MP2 (e)) considered ($T = 295$K and $p = 1$bar). Experimental values (f) taken from Ref. [58]. CN is the average coordination number calculated from the integral $4\pi\rho_n r^2 g_{OO}(r)$ up to the first minimum, where ρ_n is the average number density of Oxygen atoms.

water hydrogen bonds, and the presence of dispersion interactions. Moreover, it is known from classical simulations that, for flexible water models, the inclusion of nuclear quantum effects leads to less structured liquid and improve the density behavior. [157, 158, 159] By observing that these effects are less pronounced in D_2O than in H_2O and that the former has a molar density 1.3% higher that the latter, we expect the inclusion of this correction to reduce further the density. [159, 160] This result stands also in sharp contrast to that of BLYP, a dispersion free functional that yields 0.797 g/mL. The reason for this is that the water density depends crucially on the medium to long range part of the potential. [114, 161] Dispersion corrected density functionals, BLYP-D3, PBE-D3, and PBE0-ADMM-D3 all have densities that are much closer to experiment, albeit slightly too dense. The density obtained for PBE0-ADMM-D3 (1.023 g/mL) is of MP2 quality, but this might be in part fortuitous, as the small basis employed in the ADMM approach could influence this result. Nevertheless, it is known that the PBE0

functional improves PBE results for structural, spectroscopic and thermodynamics properties [141, 162, 163], and yields good results for water dimer interactions when compared to high level coupled cluster calculations. [164]

Computing a first principles estimate of the density is challenging, and long simulations are essential to sample fluctuations and to equilibrate the system. If a significant structural reorganization would be needed, equilibration times could easily exceed simulation times. In this respect, our RI-MP2 simulation is a 'best effort' simulation, with a length constrained by the high computational cost. The number of MC cycles is larger in the corresponding DFT simulations. However, two observations enhance our confidence in the computed MP2 value. First, the total number of *accepted* MC cycles is similar for all reported calculations. This can be attributed to the high quality of DFT presampling in the MP2 case, which results in a much higher acceptance rate ($\sim 50\%$) compared to the other simulations using presampling based on a classical force field ($15 - 17\%$). Second, those methods that remain close to the experimental density (RI-MP2 and PBE0-ADMM-D3) are likely, and observed, to require shorter equilibration periods.

Further support for the employed methodology, but also insight in the typical error bars, comes from a comparison with literature DFT results. At the BLYP level, the calculated density is 0.797 g/mL, in agreement with the results reported by McGrath and coworkers [86] (0.8 g/ml), obtained by NpT-MC, Schmidt *et al.* [79] (0.73-0.78 g/mL) calculated employing NpT-BOMD, and Wang *et al.* [161] (0.76-0.85 g/mL). In the latter case, the water density is inferred from pressure density curves obtained from series of NVT-MD simulations performed at different volumes. A larger deviation is observed when the BLYP density is compared to the value calculated by Ma and coworkers [117] (0.92 g/mL) from NpT-CPMD using a DVR basis set. In that work, the large deviation is attributed to the usage of a converged basis set that contributes to softening the structure and improving diffusivity. [165] However, even though the basis set used in this work is significantly larger than in Ref. [79] and [86], a strong basis set dependence of the density at the BLYP level is not observed. The value calculated at the BLYP-D3 level (1.066 g/mL) is in agreement with that reported by Ma *et al.* [117] (1.07

- 1.13 g/mL). When comparing our BLYP-D3 and PBE-D3 with BLYP-D2 and PBE-D2 results reported by Schmidt and coworkers [79] deviations of 7% and 11% are observed, respectively. These deviations have to be attributed to the use of the D3 correction instead of D2, or potentially the use of larger basis sets and simulation lengths in this work.

The more detailed structure of the liquid is summarized in 6.2 by the radial distribution functions for Oxygen-Oxygen ($g_{OO}(r)$) and Oxygen-Hydrogen ($g_{OH}(r)$), and quantified in 6.1. The MP2 results are in good agreement with experiment, in particular the mid and long range parts. The first minimum is relatively shallow, suggesting that a diffusive liquid is obtained. However, MD is required to obtain a precise value for the diffusion constant, as time correlation functions can not be extracted from these MC simulations. The maximum of the first peak is too pronounced, i.e. 3.12 vs 2.57. One possible explanation for this difference is that our simulations ignore nuclear quantum effects, which could well influence this property. The coordination number, which condenses the shape of the first peak into a single number, is in fair agreement with experiment 4.6 vs. 4.3. The DFT results are less satisfactory, in particular BLYP without dispersion produces a very structured liquid as quantified by the low value of the first minimum. Note that the $g_{OO}(r)$ obtained under NpT conditions is significantly different from the one obtained in the NVE ensemble (see e.g. Ref. [115]) at experimental density. In particular, a low coordination number is observed. The D3 dispersion correction leads to a $g_{OO}(r)$ that is almost featureless after the first peak. The large coordination number can be considered an artefact of the shallow minimum observed, and depends strongly on the precise location of the minimum. PBE-D3 and PBE0-ADMM-D3 give very similar $g_{OO}(r)$, as was previously observed [145] in simulations without the dispersion correction. Whereas these functionals lead to a slightly more structured liquid, they are similar to the MP2 results, in particular PBE0-ADMM-D3. The latter functional could thus be a computational expedient alternative to MP2, for example to investigate nuclear quantum effects.

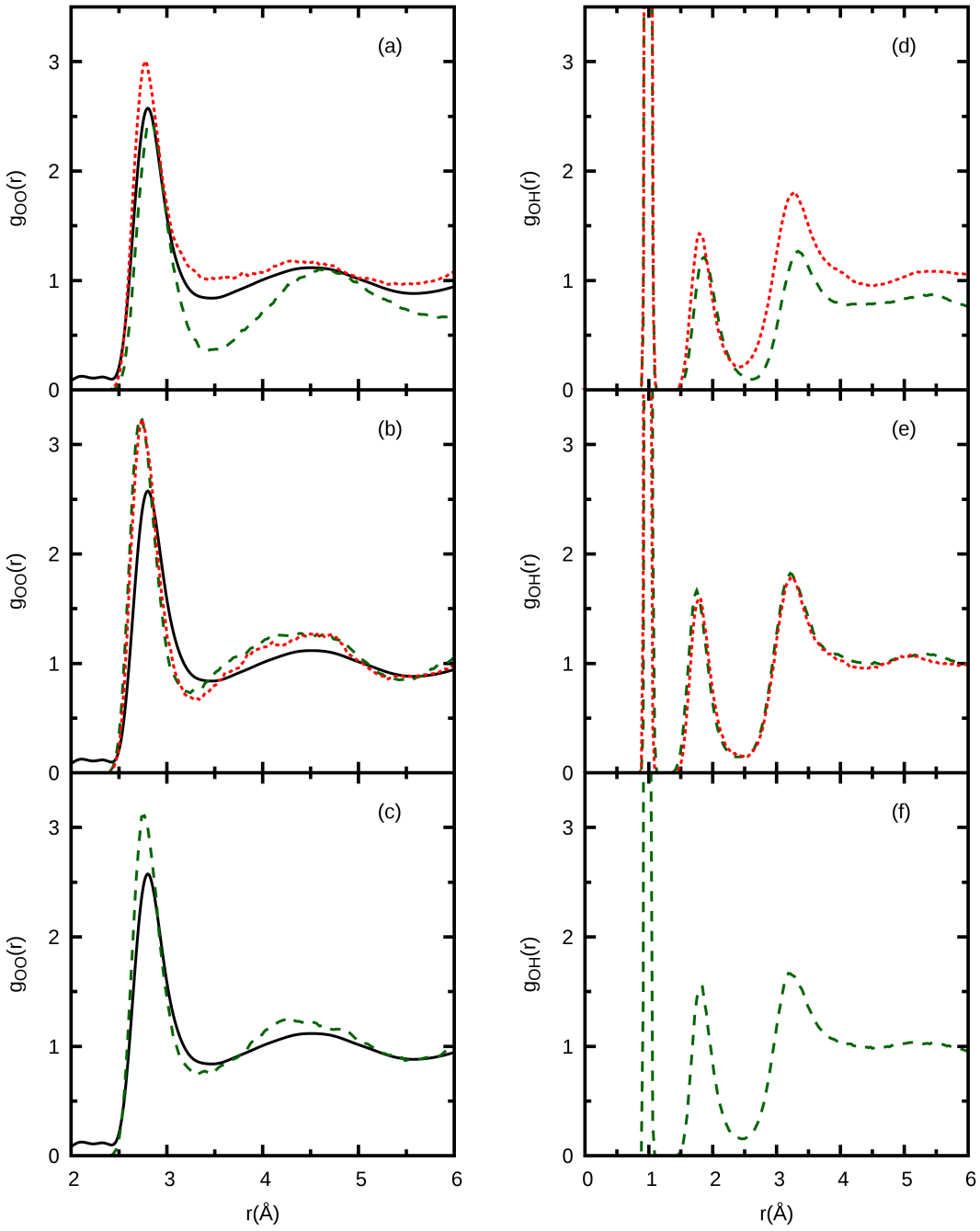


Figure 6.2: Radial distribution functions for Oxygen-Oxygen (left) and Oxygen-Hydrogen (right) distances obtained from the NpT-MC simulations at the various level of theory considered ($T = 295$K and $p = 1$bar). (a),(d) BLYP (green dashed) and BLYP-D3 (red dotted); (b),(e) PBE-D3 (green dashed) and PBE0-ADMM-D3 (red dotted); (c),(f) RI-MP2 (green dashed line). The most recent experimental Oxygen-Oxygen RDF, obtained from x-ray diffraction and taken from Ref. [58], is depicted as a solid black line.

6.4 Conclusions

Results have been reported for liquid water at ambient conditions as obtained from NpT-MC simulations at the MP2 level theory, which is free from empirical parameters. The obtained density of 1.02 g/mL is in excellent agreement with the experimental value, and the $g_{OO}(r)$ generally agrees well with the most recent experimental observations, albeit with a slightly too high first maximum. Other popular DFT methods have been tested, in particular showing good results for the water density at the PBE0-ADMM-D3 level, while being generally in agreement with previously reported work for other DFT methods. It is clear that the inclusion of dispersion interactions and an accurate description of the hydrogen bond are crucial ingredients for the correct simulation of liquid water at ambient conditions. The quality and accuracy of the obtained MP2 results paves the way to further applications of this technique for challenging questions in the field of aqueous solutions, e.g. the structures of the solvated electron and hydroxyl ion.

Chapter 7

Summary and Outlook

7.1 Summary

The first part of this thesis was dedicated to the development and implementation of the TMC algorithm into the framework of the existing molecular simulation package `CP2K`. The TMC algorithm dynamically creates a binary tree of possible Markov chain elements, assuming both possible outcomes of each MC step. As a result, the elements can be calculated simultaneously, thus the performance of a long Markov chain simulation is improved logarithmically with respect to the number of employed workers. With the help of the pre-sampling method, configurational changes can be calculated beforehand and delays in the energy calculations are prevented, which additionally reduces the overall simulation time. In the described TMC algorithm, the Markov chain is one possible way through the tree. Elements beside this Markov chain are essential for the parallel implementation, but discarded when averaging the system properties. Optimizations reduce the number of these calculated elements, concentrating the available computing power on the Markov chain elements. Thus, accurate estimates of the a priori unknown Markov chain directions, and dynamic cancellations significantly enhance the logarithmic speedup of the algorithm, which is verified and presented in this work.

Another important aspect of Monte Carlo simulations are efficient MC

moves, which significantly influence the convergence of the desired system properties with respect to the number of sampled configurations (and thus the number of calculated energies). In the context of this work, specialized moves are implemented to improve molecular simulations, especially of ice structures. Enlarged configurational changes can be obtained with the nested MC method, which at the same time drastically reduces the computational effort for the moves by employing an approximate potential. As a result, the number of exact energy calculations is reduced for obtaining system properties of a desired accuracy. For systems with large energy barriers the sampling is improved by utilizing the parallel tempering method, which is implemented with the speculative approaches of TMC. Furthermore, moves are specialized with respect to known system information e.g. the structure of ice to deploy a collective proton reordering while conserving the Bernal-Fowler ice rules. As a result the sampling of e.g. the dipole moment is remarkably improved.

The implemented algorithm was applied for calculations of water ice structures at constant temperature and constant pressure (NPT) as well as verifications at constant volume (NVT). The calculated temperature dependent dielectric constants provide better agreement to the experimental results for hybrid DFT functionals than semi-local DFT functionals. The phase transition temperatures have been determined both from a Curie-Weiss fit and from direct simulations. As a result, the PBE0-D2 model is expected to be most predictive compared to the other investigated functionals. Furthermore, bulk liquid water calculations at the MP2 level theory are presented, resulting in a density of 1.02 g/mL, which is in excellent agreement with the experimental value. Where the $g_{OO}(r)$ generally agrees well with the most experimental observations, albeit with a slightly too high first maximum. In conclusion, it is shown that the steady increase in computing power and the development of the novel TMC algorithm have made the computational intensive sampling of accurate molecular systems at the DFT level of theory possible.

7.2 Outlook

An improved prediction of the Markov chain directions would lead to better scaling in performance (with respect to the number of workers). A perfect predictability of the Markov chain direction would lead to an enhanced scaling behavior getting closer and closer to the linear scaling.

With the increasing sizes of high-performance computing systems, faults are a norm rather than an exception. Usually, in case of compute node failures the whole simulation stops. Nowadays the simulations are continued utilizing checkpoint-restart techniques (also in TMC), but the user has to deal with possibly corrupted data. These node failures could be easily handled within the dynamic task scheduling of the TMC algorithm using the framework provided by `MPI` 3.0[10, 166]. Unavailable nodes can be traced and excluded from the work flow, unless the master process is effected. The simulation could keep running with a decreased number of processors. A dynamic reallocation of nodes would be a suitable extension to this algorithm.

Furthermore, parallel tempering has been implemented and tested with respect to consistency and correctness, but investigations of the efficiency and scaling with respect to utilized processors should be done in more detail. Additionally, the set of implemented moves should be enhanced (e.g. adding the exchange of two atoms of different kind) to improve the simulations of different kind of systems.

Moreover, the speculative approach of TMC can be applied for optimization algorithms. For example, the Gaussian adaptation algorithm[167], which creates samples within a space defined by a mean point, a radius and an covariance matrix. After an acceptance check, parameters are adapted, but independent of the value of the calculated energy (objective function). This is an intrinsically serial procedure, but both possible outcomes of the acceptance check can be assumed (similar to the TMC algorithm) to create possibly new samples and calculate them in parallel. Only few changes would be necessary to incorporate the Gaussian adaptation algorithm in TMC. Similar to this, also other optimization algorithm like covariance matrix adaptation Evolution Strategy (CMA-ES)[168], simulated annealing[169] could employ

the speculative approach, predicting next steps and calculate them in parallel.

References

[1] T. Ziegler, "Approximate density functional theory as a practical tool in molecular energetics and dynamics," *Chem. Rev.*, vol. 91, no. 5, pp. 651–667, 1991.

[2] N. M. Harrison, "An introduction to density functional theory," in *Comp. Mater. Sci.*, pp. 45–70, 2003.

[3] N. Metropolis and A. Rosenbluth, "Equation of state calculations by fast computing machines," *J. Chem. Phys.*, vol. 21, no. 6, pp. 1087–1092, 1953.

[4] D. J. Wilkinson, "Bayesian methods in bioinformatics and computational systems biology.," *Brief. Bioinform.*, vol. 8, pp. 109–16, Mar. 2007.

[5] M. Bonamente, M. K. Joy, J. E. Carlstrom, E. D. Reese, and S. J. LaRoque, "Markov Chain Monte Carlo joint analysis of Chandra X-ray imaging spectroscopy and Sunyaev-Zeldovich Effect data," *Astrophys. J.*, vol. 614, no. 1, pp. 1–24, 2004.

[6] M. Johannes and N. Polson, "MCMC methods for financial econometrics," in *The Handbook of Financial Econometrics*, no. Mcmc, pp. 1–76, 2002.

[7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, "Equation of State Calculations by Fast Computing Machines," *J. Chem. Phys.*, vol. 21, no. 6, p. 1087, 1953.

[8] D. Frenkel and B. Smit, eds., *Understanding molecular simulation: from algorithms to applications*, vol. 50 of *Computational science*. Academic Press, 2002.

[9] V. Manousiouthakis, "Strict detailed balance is unnecessary in Monte Carlo simulation," *J. Chem. Phys.*, vol. 110, no. 6, pp. 6–9, 1999.

[10] E. Lusk, S. Huss, B. Saphir, and M. Snir, *MPI: A message-passing interface standard*. University of Tennessee, Knoxville, 2012.

[11] J. Rosenthal, "Parallel computing and Monte Carlo algorithms," *Far East Journal of Theoretical Statistics*, vol. 4, pp. 1–27, 2000.

[12] K. Esselink, L. D. J. C. Loyens, and B. Smit, "Parallel Monte Carlo simulations," *Phys. Rev. E*, vol. 51, no. 2, pp. 1560–1568, 1995.

[13] M. Lingenheil, R. Denschlag, G. Mathias, and P. Tavan, "Efficiency of exchange schemes in replica exchange," *Chem. Phys. Lett.*, vol. 478, pp. 80–84, Aug. 2009.

[14] J. M. R. Byrd, S. a. Jarvis, and A. H. Bhalerao, "On the parallelisation of MCMC by speculative chain execution," *IPDPSW*, pp. 1–8, Apr. 2010.

[15] J. Byrd, S. Jarvis, and A. Bhalerao, "Speculative Moves: Multithreading Markov Chain Monte Carlo Programs," *High-Performance Medical Image Computing and Computer Aided Intervention*, p. 12, 2008.

[16] J. M. R. Byrd, S. A. Jarvis, and A. H. Bhalerao, "Reducing the run-time of MCMC programs by multithreading on SMP architectures," *IPDPS*, pp. 1–8, 2008.

[17] A. Brockwell, "Parallel Markov chain Monte Carlo simulation by prefetching," *J. Comput. Graph. Stat.*, vol. 15, no. 1, pp. 1–18, 2006.

[18] A. Gelman, G. Roberts, and W. Gilks, "Efficient metropolis jumping hules," *Bayesian statistics*, vol. 5, pp. 599–607, 1996.

[19] S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth, "Hybrid Monte Carlo," *Phys. Lett. B*, vol. 195, no. 2, pp. 216–222, 1987.

[20] L. D. Gelb, "Monte Carlo simulations using sampling from an approximate potential," *J. Chem. Phys.*, vol. 118, no. 17, pp. 7747–7750, 2003.

[21] "The CP2K developer group, CP2K is freely available from: http://www.cp2k.org/," 2013.

[22] I. Bethune, "Improving the Performance of CP2K on the Cray XT," in *CUG 2010 Proceedings*, pp. 1–7, The Cray User Group, 2010.

[23] R. Rabenseifner, G. Hager, and G. Jost, "Hybrid MPI/OpenMP Parallel Programming on Clusters of Multi-Core SMP Nodes," in *PDP Proceedings*, no. c, pp. 427–436, Ieee, 2009.

[24] R. Iftimie and D. Salahub, "Using a classical potential as an efficient importance function for sampling from an ab initio potential," *Chem. Phys.*, vol. 113, no. 12, pp. 4852–4862, 2000.

[25] J. D. Coe, T. D. Sewell, and M. S. Shaw, "Optimal sampling efficiency in Monte Carlo simulation with an approximate potential.," *J. Chem. Phys.*, vol. 130, p. 164104, Apr. 2009.

[26] L. Gelb, "Monte Carlo strategies for first-principles simulations of elemental systems," in *XSEDE12*, p. 7, 2012.

[27] D. Earl and M. Deem, "Parallel tempering: Theory, applications, and new perspectives," *Phys. Chem. Chem. Phys.*, vol. 7, pp. 3910–3916, 2005.

[28] R. Swendsen and J. Wang, "Replica Monte Carlo simulation of spin glasses," *Phys. Rev. Lett.*, vol. 57, pp. 2607–2609, Nov. 1986.

[29] G. Bussi, D. Fisica, and R. Emilia, "A simple asynchronous replica-exchange implementation," *Processing*, pp. 1–3, 2008.

[30] A. N. Tharrington and K. D. Jordan, "Parallel-Tempering Monte Carlo Study of (H 2 O) n = 6 - 9," *J. Phys. Chem. A*, vol. 107, pp. 7380–7389, Sept. 2003.

[31] J. D. Bernal and R. H. Fowler, "A Theory of Water and Ionic Solution, with Particular Reference to Hydrogen and Hydroxyl Ions," *J. Chem. Phys.*, vol. 1, no. 8, pp. 515–548, 1933.

[32] G. P. Johari and E. Whalley, "The dielectric properties of ice Ih in the range 272-133 K," *J. Chem. Phys.*, vol. 75, no. 3, pp. 1333–1340, 1981.

[33] S. W. Rick, "Simulations of proton order and disorder in ice Ih.," *J. Chem. Phys.*, vol. 122, p. 094504, Mar. 2005.

[34] S. W. Rick and a. D. J. Haymet, "Dielectric constant and proton order and disorder in ice Ih: Monte Carlo computer simulations," *J. Chem. Phys.*, vol. 118, no. 20, pp. 9291–9296, 2003.

[35] G. Barkema and M. Newman, "Monte Carlo simulation of ice models," *Phys. Rev. E*, vol. 57, pp. 1155–1166, Jan. 1998.

[36] G. T. Barkema and J. D. Boer, "Properties of a statistical model of ice at low temperatures," *J. Chem. Phys.*, vol. 99, no. 3, p. 2059, 1993.

[37] P. L. Ecuyer, R. Simard, E. J. Chen, and W. D. Kelton, "An Object-oriented Random-number Package with many long streams and substreams," *Operations Research*, vol. 50, no. 6, pp. 1073–1075, 2002.

[38] P. Mutzel, M. Jünger, and S. Leipert, eds., *Graph Drawing*, vol. 2265 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, Feb. 2002.

[39] M. Schönherr, B. Slater, J. Hutter, and J. VandeVondele, "Dielectric Properties of Water Ice, the Ice Ih/XI Phase Transition and an Assessment of DFT," *J. Phys. Chem. B*, vol. accepted, 2013.

[40] M. Faraday, "On Certain Conditions of Freezing Water," *Athemaeum*, vol. 1181, pp. 640–641, 1850.

[41] M. B. Watkins, D. Pan, E. G. Wange, A. Michaelides, J. VandeVondele, and B. Slater, "Large Variation of Vacancy Formation Energies in the Surface of Crystalline Ice," *Nat. Mater.*, vol. 10, pp. 794–798, 2011.

[42] C. G. Salzmann, P. G. Radaelli, B. Slater, and J. L. Finney, "The polymorphism of ice: five unresolved questions.," *Phys. Chem. Chem. Phys.*, vol. 13, pp. 18468–80, Nov. 2011.

[43] L. Pauling, "The Structure and Entropy of Ice and of Other Crystals With Some Randomness of Atomic Arrangement," *J. Am. Chem. Soc.*, vol. 57, no. 1920, pp. 2680–2684, 1935.

[44] W. F. Giauque and J. W. Stout, "The Entropy of Water and the Third Law of Thermodynamics. The Heat Capacity of Ice from 15 to 273K," vol. 58, pp. 1144–1150, 1936.

[45] T. K. Hirsch and L. Ojama, "Quantum-Chemical and Force-Field Investigations of Ice Ih : Computation of Proton-Ordered Structures and Prediction of Their Lattice Energies," *J. Phys. Chem. B*, vol. 108, no. 40, pp. 15856–15864, 2004.

[46] G. P. Johari and S. J. Jones, "The orientation polarization in hexagonal ice parallel and perpendicular to the c-axis," *J. Glacialology*, vol. 21, pp. 259–276, 1978.

[47] I. Minagawa, "Ferroelectric Phase Transition and Anisotropy of Dielectric Constant in Ice Ih," *J. Phys. Soc. Jpn.*, vol. 50, pp. 3669–3676, 1981.

[48] J. L. Aragones, L. G. MacDowell, and C. Vega, "Dielectric constant of ices and water: a lesson about water interactions.," *J. Phys. Chem. A*, vol. 115, pp. 5745–5758, June 2011.

[49] F. Humbel, F. Jona, and P. Scherrer, "Anisotropie der Dielektrizitätskonstante des Eises," *Helv. Phys. Acta*, vol. 26, no. 1, pp. 17–32, 1953.

[50] S. Kawada, "Dielectric anisotropy in ice Ih," *J. Phys. Soc. Japan*, vol. 44, no. 6, pp. 1881–1886, 1978.

[51] S. Kawada, "Dielectric dispersion and phase transition of KOH doped ice," *J. Phys. Soc. Japan*, vol. 32, p. 1442, 1972.

[52] Y. Tajima, T. Matsuo, and H. Suga, "Phase transition in KOH-doped hexagonal ice," *Nature*, vol. 299, pp. 810–812, 1982.

[53] S. Kawada, "Acceleration Of Dielectric Relaxation By KOH-Doping And Phase Transition In Ice Ih," *J. Phys. Chem. Solids*, vol. 50, no. 11, pp. 1177–1184, 1989.

[54] M. Oguro and R. Whitworth, "Dielectric observations of the transformation of single crystals of KOH-doped ice Ih to ice XI," *J. Phys. Chem. Solids*, vol. 52, pp. 401–403, Jan. 1991.

[55] S. M. Jackson and R. W. Whitworth, "Evidence for ferroelectric ordering of ice Ih," *J. Chem. Phys.*, vol. 103, no. 17, pp. 7647–7648, 1995.

[56] S. Jackson and V. Nield, "Single-crystal neutron diffraction studies of the structure of ice XI," *J. Phys. Chem. B*, vol. 5647, no. 96, pp. 6142–6145, 1997.

[57] A. J. Leadbetter, R. C. Ward, J. W. Clark, P. a. Tucker, T. Matsuo, and H. Suga, "The equilibrium low-temperature structure of ice," *J. Chem. Phys.*, vol. 82, no. 1, p. 424, 1985.

[58] L. B. Skinner, C. Huang, D. Schlesinger, L. G. M. Pettersson, A. Nilsson, and C. J. Benmore, "Benchmark oxygen-oxygen pair-distribution function of ambient water from x-ray diffraction measurements with a wide Q-range," *J. Chem. Phys.*, vol. 138, no. 7, p. 74506, 2013.

[59] G. E. Lindberg and F. Wang, "Efficient sampling of ice structures by electrostatic switching.," *J. Phys. Chem. B*, vol. 112, pp. 6436–41, May 2008.

[60] L. G. MacDowell and C. Vega, "Dielectric constant of ice Ih and ice V: a computer simulation study.," *J. Phys. Chem. B*, vol. 114, pp. 6089–6098, May 2010.

[61] G. a. Tribello and B. Slater, "Proton ordering energetics in ice phases," *Chem. Phys. Lett.*, vol. 425, pp. 246–250, July 2006.

[62] S. Singer, J.-L. Kuo, T. Hirsch, C. Knight, L. Ojamäe, and M. Klein, "Hydrogen-Bond Topology and the Ice VII/VIII and Ice Ih/XI Proton-Ordering Phase Transitions," *Phys. Rev. Lett.*, vol. 94, p. 135701, Apr. 2005.

[63] S. Casassa, M. Calatayud, K. Doll, C. Minot, and C. Pisani, "Proton ordered cubic and hexagonal periodic models of ordinary ice," *Chem. Phys. Lett.*, vol. 409, pp. 110–117, June 2005.

[64] F. Labat, C. Pouchan, C. Adamo, and G. E. Scuseria, "Role of Nonlocal Exchange in Molecular Crystals: The Case of Two Proton-ordered Phases of Ice," *J. Comput. Chem.*, vol. 32, no. 10, pp. 2177–2185, 2011.

[65] C. Knight, S. Singer, J.-L. Kuo, T. Hirsch, L. Ojamäe, and M. Klein, "Hydrogen bond topology and the ice VII/VIII and Ih/XI proton ordering phase transitions," *Phys. Rev. E*, vol. 73, pp. 1–14, May 2006.

[66] A. J. Rusnak, E. R. Pinnick, C. E. Calderon, and F. Wang, "Static dielectric constants and molecular dipole distributions of liquid water and ice-Ih investigated by the PAW-PBE exchange-correlation functional.," *J. Chem. Phys.*, vol. 137, p. 034510, July 2012.

[67] R. King-Smith and D. Vanderbilt, "Theory of polarization of crystalline solids," *Phys. Rev. B*, vol. 47, no. 3, pp. 1651–1654, 1993.

[68] N. a. Spaldin, "A beginner's guide to the modern theory of polarization," *J. Solid State Chem.*, vol. 195, pp. 2–10, Nov. 2012.

[69] M. Sharma, R. Resta, and R. Car, "Dipolar Correlations and the Dielectric Permittivity of Water," *Phys. Rev. Lett.*, vol. 98, p. 247401, June 2007.

[70] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach," *Comp. Phys. Comm.*, vol. 167, pp. 103–128, Apr. 2005.

[71] I. McDonald, "NpT-ensemble Monte Carlo calculations for binary liquid mixtures," *Mol. Phys.*, vol. 23, no. 1, pp. 41–58, 1972.

[72] J. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct. 1996.

[73] A. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, no. 6, pp. 3098–3100, 1988.

[74] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a function of the electron density," *Phys. Rev. B*, vol. 37, no. 2, pp. 785–789, 1988.

[75] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," *J. Chem. Phys.*, vol. 110, no. 13, p. 6158, 1999.

[76] A. D. Becke, "Density-functional thermochemistry. III. The role of exact exchange," *J. Chem. Phys.*, vol. 98, no. 7, pp. 5648–5652, 1993.

[77] S. H. Vosko, L. Wilk, and M. Nusair, "Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis," *Can. J. Phys.*, vol. 58, no. 8, pp. 1200–1211, 1980.

[78] S. Grimme, "Semiempirical GGA-type density functional constructed with a longrange dispersion correction," *J. Comput. Chem.*, vol. 27, no. 15, pp. 1787–1799, 2006.

[79] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy, "Isobaric-isothermal molecular dynamics simulations utilizing density functional theory: an assessment of the structure and density of water at near-ambient conditions.," *J. Phys. Chem. B*, vol. 113, pp. 11959–64, Sept. 2009.

[80] S. Goedecker, M. Teter, and J. Hutter, "Seperable dual-space Gaussian pseudopotentials," *Phys. Rev. B*, vol. 54, no. 3, pp. 1703–1710, 1996.

[81] M. Guidon, F. Schiffmann, J. Hutter, and J. VandeVondele, "Ab initio molecular dynamics using hybrid density functionals.," *J. Chem. Phys.*, vol. 128, p. 214104, June 2008.

[82] M. Guidon, J. Hutter, and J. VandeVondele, "Robust Periodic Hartree Fock Exchange for Large-Scale Simulations Using Gaussian Basis Sets," *J. Chem. Theory Comput.*, vol. 5, no. 11, pp. 3010–3021, 2009.

[83] M. Guidon, *High Performance Hartree-Fock Exchange for Large and Condensed Phase Systems*. PhD thesis, 2010.

[84] S. W. de Leeuw, J. W. Perram, and E. R. Smith, "Simulation of Electrostatic Systems in Periodic Boundary Conditions. I. Lattice Sums and Dielectric Constants," *Proc. R. Soc. Lond. A*, vol. 373, pp. 27–56, Oct. 1980.

[85] M. Stengel, N. Spaldin, and D. Vanderbilt, "Electric Displacement as the Fundamental Variable in Electronic-structure Calculations," *Nature Phys.*, vol. 5, pp. 304–308, 2009.

[86] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack, "Isobaric-isothermal Monte Carlo Simulations from First Principles: Application to Liquid Water at Ambient Conditions," *ChemPhysChem*, vol. 6, pp. 1894–901, Sept. 2005.

[87] S. Izvekov and J. M. J. Swanson, "Using force-matching to reveal essential differences between density functionals in ab initio molecular dynamics simulations," *J.Chem. Phys.*, vol. 134, no. 19, p. 194109, 2011.

[88] K. Röttger, A. Endriss, J. Ihringer, S. Doyle, and W. F. Kuhs, "Lattice constants and thermal expansion of H2O and D2O ice I h between 10 and 265 K," *Acta Cryst. B*, vol. 50, pp. 644–648, Dec. 1994.

[89] R. Howe and R. W. Whitworth, "A determination of the crystal structure of ice XI," *J. Chem. Phys.*, vol. 90, no. 8, pp. 4450–4453, 1989.

[90] S. McDowell, R. Amos, and N. Handy, "Molecular polarisabilities-a comparison of density functional theory with standard ab initio methods," *Chem. Phys. Lett.*, vol. 235, pp. 1–4, 1995.

[91] M. Del Ben, M. Schönherr, J. Hutter, and J. VandeVondele, "Bulk Liquid Water at Ambient Temperature and Pressure From MP2 Theory," *J. Phys. Chem. Lett.*, vol. 4, pp. 3753–3759, 2013.

[92] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

[93] B. Guillot, "A reappraisal of what we have learnt during three decades of computer simulations on water," *J. Mol. Liq.*, vol. 101, no. 1-3, pp. 219–260, 2002.

[94] W. L. Jorgensen and J. Tirado-Rives, "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 19, pp. 6665–6670, 2005.

[95] R. Bukowski, K. Szalewicz, G. C. Groenenboom, and A. van der Avoird, "Predictions of the Properties of Water from First Principles," *Science*, vol. 315, no. 5816, pp. 1249–1252, 2007.

[96] A. G. Donchev, N. G. Galkin, A. A. Illarionov, O. V. Khoruzhii, M. A. Olevanov, V. D. Ozrin, M. V. Subbotin, and V. I. Tarasov, "Water properties from first principles: Simulations by a general-purpose quantum mechanical polarizable force field," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 103, no. 23, pp. 8613–8617, 2006.

[97] M. W. Mahoney and W. L. Jorgensen, "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions," *J. Chem. Phys.*, vol. 112, no. 20, pp. 8910–8922, 2000.

[98] S. W. Rick, S. J. Stuart, and B. J. Berne, "Dynamical fluctuating charge force fields: Application to liquid water," *J. Chem. Phys.*, vol. 101, no. 7, pp. 6141–6156, 1994.

[99] P. Paricaud, M. Predota, A. A. Chialvo, and P. T. Cummings, "From dimer to condensed phases at extreme conditions: Accurate predictions of the properties of water by a Gaussian charge polarizable model," *J. Chem. Phys.*, vol. 122, no. 24, p. 244511, 2005.

[100] B. Chen, J. Xing, and J. I. Siepmann, "Development of Polarizable Water Force Fields for Phase Equilibrium Calculations," *J. Phys. Chem. B*, vol. 104, no. 10, pp. 2391–2401, 2000.

[101] T. M. Truskett and K. A. Dill, "A Simple Statistical Mechanical Model of Water," *J. Phys. Chem. B*, vol. 106, no. 45, pp. 11829–11842, 2002.

[102] J. R. Errington and P. G. Debenedetti, "Relationship between structural order and the anomalies of liquid water," *Nature*, vol. 409, pp. 318–321, 2001.

[103] H. Stanley, S. V. Buldyrev, and N. Giovambattista, "Static heterogeneities in liquid water," *Physica A*, vol. 342, no. 1-2, pp. 40–47, 2004.

[104] A. Nilsson and L. G. M. Pettersson, "Perspective on the structure of liquid water," *Chem. Phys.*, vol. 389, no. 1-3, pp. 1–34, 2011.

[105] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, M. Krack, and M. Parrinello, "Liquid Water from First Principles: Investigation of Different Sampling Approaches," *J. Phys. Chem. B*, vol. 108, no. 34, pp. 12990–12998, 2004.

[106] K. Laasonen, M. Sprik, M. Parrinello, and R. Car, ""Ab Initio" Liquid Water," *J. Chem. Phys.*, vol. 99, no. 11, pp. 9080–9089, 1993.

[107] P. L. Silvestrelli and M. Parrinello, "Structural, electronic, and bonding properties of liquid water from first principles," *J. Chem. Phys.*, vol. 111, no. 8, pp. 3572–3580, 1999.

[108] S. Izvekov and G. A. Voth, "Car–Parrinello molecular dynamics simulation of liquid water: New results," *J. Chem. Phys.*, vol. 116, no. 23, pp. 10372–10376, 2002.

[109] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. A. Näslund, T. K. Hirsch, L. Ojamäe, P. Glatzel, L. G. M. Pettersson, and A. Nilsson, "The Structure of the First Coordination Shell in Liquid Water," *Science*, vol. 304, no. 5673, pp. 995–999, 2004.

[110] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water," *J. Chem. Phys.*, vol. 120, no. 1, pp. 300–311, 2004.

[111] T. D. Kühne, M. Krack, and M. Parrinello, "Static and Dynamical Properties of Liquid Water from First Principles by a Novel Car–Parrinello–like Approach," *J. Chem. Theory Comput.*, vol. 5, no. 2, pp. 235–241, 2009.

[112] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger, "Structure and Dynamics of Liquid Water from ab Initio Molecular Dynamics–Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections," *J. Chem. Theory Comput.*, vol. 8, no. 10, pp. 3902–3910, 2012.

[113] E. Schwegler, J. C. Grossman, F. Gygi, and G. Galli, "Towards an assessment of the accuracy of density functional theory for first principles simulations of water. II," *J. Chem. Phys.*, vol. 121, no. 11, pp. 5400–5409, 2004.

[114] R. Jonchiere, A. P. Seitsonen, G. Ferlat, A. M. Saitta, and R. Vuilleumier, "Van der Waals effects in ab initio water at ambient and supercritical conditions," *J. Chem. Phys.*, vol. 135, no. 15, p. 154503, 2011.

[115] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello, "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water," *J. Chem. Phys.*, vol. 122, no. 1, p. 14515, 2005.

[116] A. Mø gelhøj, A. K. Kelkkanen, K. T. Wikfeldt, J. Schiø tz, J. J. r. Mortensen, L. G. M. Pettersson, B. I. Lundqvist, K. W. Jacobsen, A. Nilsson, and J. K. Nø rskov, "Ab Initio van der Waals Interactions in Simulations of Water Alter Structure from Mainly Tetrahedral to High-Density-Like," *J. Phys. Chem. B*, vol. 115, no. 48, pp. 14149–14160, 2011.

[117] Z. Ma, Y. Zhang, and M. E. Tuckerman, "Ab initio molecular dynamics study of water at constant pressure using converged basis sets and empirical dispersion corrections," *J. Chem. Phys.*, vol. 137, no. 4, p. 44506, 2012.

[118] S. Nosé and M. L. Klein, "Constant pressure molecular dynamics for molecular systems," *Mol. Phys.*, vol. 50, no. 5, pp. 1055–1076, 1983.

[119] R. Iftimie, D. Salahub, D. Wei, and J. Schofield, "Using a classical potential as an efficient importance function for sampling from an ab initio potential," *J. Chem. Phys.*, vol. 113, no. 12, pp. 4852–4862, 2000.

[120] B. Hetenyi, K. Bernacki, and B. J. Berne, "Multiple "time step" Monte Carlo," *J. Chem. Phys.*, vol. 117, no. 18, pp. 8203–8207, 2002.

[121] L. D. Gelb, "Monte Carlo simulations using sampling from an approximate potential," *J. Chem. Phys.*, vol. 118, no. 17, pp. 7747–7750, 2003.

[122] B. Mehlig, D. W. Heermann, and B. M. Forrest, "Hybrid Monte Carlo method for condensed-matter systems," *Phys. Rev. B*, vol. 45, pp. 679–685, Jan. 1992.

[123] A. D. Becke, "Density-functional exchange-energy approximation with correct asymptotic behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, Sept. 1988.

[124] C. Lee, W. Yang, and R. G. Parr, "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density," *Phys. Rev. B*, vol. 37, pp. 785–789, Jan. 1988.

[125] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized Gradient Approximation Made Simple," *Phys. Rev. Lett.*, vol. 77, pp. 3865–3868, Oct. 1996.

[126] S. Grimme, "Accurate description of van der Waals complexes by density functional theory including empirical corrections," *J. Comput. Chem.*, vol. 25, no. 12, pp. 1463–1473, 2004.

[127] S. Grimme, "Semiempirical GGA-type density functional constructed with a long-range dispersion correction," *J. Comput. Chem.*, vol. 27, no. 15, pp. 1787–1799, 2006.

[128] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, "A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu," *J. Chem. Phys.*, vol. 132, no. 15, p. 154104, 2010.

[129] C. Mø ller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," *Phys. Rev.*, vol. 46, pp. 618–622, Oct. 1934.

[130] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*. New York: McGraw Hill, 1982.

[131] S. S. Xantheas, C. J. Burnham, and R. J. Harrison, "Development of transferable interaction models for water. II. Accurate energetics of the first few water clusters from first principles," *J. Chem. Phys.*, vol. 116, no. 4, pp. 1493–1499, 2002.

[132] S. S. Xantheas and E. Apra, "The binding energies of the D_{2d} and S_{4} water octamer isomers: High–level electronic structure and empirical potential results," *J. Chem. Phys.*, vol. 120, no. 2, pp. 823–828, 2004.

[133] I. Shin, M. Park, S. K. Min, E. C. Lee, S. B. Suh, and K. S. Kim, "Structure and spectral features of Hˆ+(H_2O)_7: Eigen versus Zundel forms," *J. Chem. Phys.*, vol. 125, no. 23, p. 234305, 2006.

[134] E. E. Dahlke, R. M. Olson, H. R. Leverentz, and D. G. Truhlar, "Assessment of the Accuracy of Density Functionals for Prediction of Relative Energies and Geometries of Low-Lying Isomers of Water Hexamers," *J. Phys. Chem. A*, vol. 112, no. 17, pp. 3976–3984, 2008.

[135] L. Goerigk and S. Grimme, "A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions," *Phys. Chem. Chem. Phys.*, vol. 13, no. 14, pp. 6670–6688, 2011.

[136] A. Grüneis, M. Marsman, and G. Kresse, "Second-order Møller–Plesset perturbation theory applied to extended systems. II. Structural and energetic properties," *J. Chem. Phys.*, vol. 133, no. 7, p. 74107, 2010.

[137] A. Erba, L. Maschio, S. Salustro, and S. Casassa, "A post-Hartree–Fock study of pressure-induced phase transitions in solid nitrogen: The case of the

alpha, gamma, and epsilon low-pressure phases," *J. Chem. Phys.*, vol. 134, no. 7, p. 74502, 2011.

[138] F. Goltl, A. Gruneis, T. Bucko, and J. Hafner, "Van der Waals interactions between hydrocarbon molecules and zeolites: Periodic calculations at different levels of theory, from density functional theory to the random phase approximation and Møller-Plesset perturbation theory," *J. Chem. Phys.*, vol. 137, no. 11, p. 114111, 2012.

[139] M. Del Ben, J. Hutter, and J. VandeVondele, "Second-Order Møller–Plesset Perturbation Theory in the Condensed Phase: An Efficient and Massively Parallel Gaussian and Plane Waves Approach," *J. Chem. Theory Comput.*, vol. 8, no. 11, pp. 4177–4188, 2012.

[140] M. Del Ben, J. Hutter, and J. VandeVondele, "Electron Correlation in the Condensed Phase from a Resolution of Identity Approach Based on the Gaussian and Plane Waves Scheme," *J. Chem. Theory Comput.*, vol. 9, no. 6, pp. 2654–2671, 2013.

[141] C. Adamo and V. Barone, "Toward reliable density functional methods without adjustable parameters: The PBE0 model," *J. Chem. Phys.*, vol. 110, no. 13, pp. 6158–6170, 1999.

[142] G. Lippert, J. Hutter, and M. Parrinello, "A hybrid Gaussian and plane wave density functional scheme," *Mol. Phys.*, vol. 92, no. 3, pp. 477–488, 1997.

[143] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing, and J. Hutter, "Quickstep: Fast and accurate density functional calculations using a mixed Gaussian and plane waves approach," *Comput. Phys. Commun.*, vol. 167, no. 2, pp. 103–128, 2005.

[144] S. Goedecker, M. Teter, and J. Hutter, "Separable dual-space Gaussian pseudopotentials," *Phys. Rev. B*, vol. 54, pp. 1703–1710, July 1996.

[145] M. Guidon, F. Schiffmann, J. Hutter, and J. VandeVondele, "Ab initio molecular dynamics using hybrid density functionals," *J. Chem. Phys.*, vol. 128, p. 214104, 2008.

[146] J. Paier, C. V. Diaconu, G. E. Scuseria, M. Guidon, J. VandeVondele, and J. Hutter, "Accurate Hartree-Fock energy of extended systems using large Gaussian basis sets," *Phys. Rev. B*, vol. 80, p. 174114, Nov. 2009.

[147] M. Guidon, J. Hutter, and J. VandeVondele, "Auxiliary Density Matrix Methods for Hartree–Fock Exchange Calculations," *J. Chem. Theory Comput.*, vol. 6, no. 8, pp. 2348–2364, 2010.

[148] J. L. Whitten, "Coulombic potential energy integrals and approximations," *J. Chem. Phys.*, vol. 58, no. 10, pp. 4496–4501, 1973.

[149] B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, "On some approximations in applications of X alpha theory," *J. Chem. Phys.*, vol. 71, no. 8, pp. 3396–3402, 1979.

[150] F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, "RI-MP2: optimized auxiliary basis sets and demonstration of efficiency," *Chem. Phys. Lett.*, vol. 294, pp. 143–152, Sept. 1998.

[151] O. Vahtras, J. Almlöf, and M. W. Feyereisen, "Integral approximations for LCAO-SCF calculations," *Chem. Phys. Lett.*, vol. 213, no. 56, pp. 514–518, 1993.

[152] T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," *J. Chem. Phys.*, vol. 90, no. 2, pp. 1007–1023, 1989.

[153] D. E. Woon and T. H. Dunning, "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon," *J. Chem. Phys.*, vol. 98, no. 2, pp. 1358–1371, 1993.

[154] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, and C. J. Mundy, "Vapor-liquid equilibria of water from first principles: comparison of density functionals and basis sets," *Mol. Phys.*, vol. 104, no. 22-24, pp. 3619–3626, 2006.

[155] S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, "Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching," *J. Chem. Phys.*, vol. 120, no. 23, pp. 10896–10913, 2004.

[156] E. E. Dahlke and D. G. Truhlar, "Improved Density Functionals for Water," *J. Phys. Chem. B*, vol. 109, no. 33, pp. 15677–15683, 2005.

[157] R. A. Kuharski and P. J. Rossky, "A quantum mechanical study of structure in liquid H_2O and D_2O," *J. Chem. Phys.*, vol. 82, no. 11, pp. 5164–5177, 1985.

[158] J. Lobaugh and G. A. Voth, "A quantum model for water: Equilibrium and dynamical properties," *J. Chem. Phys.*, vol. 106, no. 6, pp. 2400–2410, 1997.

[159] M. W. Mahoney and W. L. Jorgensen, "Quantum, intramolecular flexibility, and polarizability effects on the reproduction of the density anomaly of liquid water by simple potential functions," *J. Chem. Phys.*, vol. 115, no. 23, pp. 10758–10768, 2001.

[160] F. Paesani, S. Iuchi, and G. A. Voth, "Quantum effects in liquid water from an ab initio-based polarizable force field," *J. Chem. Phys.*, vol. 127, no. 7, p. 74506, 2007.

[161] J. Wang, G. Román-Pérez, J. M. Soler, E. Artacho, and M.-V. Fernández-Serra, "Density, structure, and dynamics of water: The effect of van der Waals interactions," *J. Chem. Phys.*, vol. 134, no. 2, p. 24516, 2011.

[162] C. Zhang, D. Donadio, F. Gygi, and G. Galli, "First Principles Simulations of the Infrared Spectrum of Liquid Water Using Hybrid Density Functionals," *J. Chem. Theory Comput.*, vol. 7, no. 5, pp. 1443–1449, 2011.

[163] B. Santra, A. Michaelides, and M. Scheffler, "On the accuracy of density-functional theory exchange-correlation functionals for H bonds in small water clusters: Benchmarks approaching the complete basis set limit," *J. Chem. Phys.*, vol. 127, no. 18, p. 184104, 2007.

[164] B. Santra, A. Michaelides, and M. Scheffler, "Coupled cluster benchmarks of water monomers and dimers extracted from density-functional theory liquid water: The importance of monomer deformations," *J. Chem. Phys.*, vol. 131, no. 12, p. 124509, 2009.

[165] H.-S. Lee and M. E. Tuckerman, "Structure of liquid water at ambient temperature from ab initio molecular dynamics performed in the complete basis set limit," *J. Chem. Phys.*, vol. 125, no. 15, p. 154507, 2006.

[166] W. Gropp and E. Lusk, "Fault Tolerance in MPI Programs," *IJHPCA*, vol. 18, pp. 363–372, 2002.

[167] C. L. Muller and I. F. Sbalzarini, "Gaussian Adaptation as a unifying framework for continuous black-box optimization and adaptive Monte Carlo sampling," in *IEEE Congress on Evolutionary Computation*, pp. 1–8, Ieee, July 2010.

[168] C. Igel and N. Hansen, "A Computational Efficient Covariance Matrix Update and a (1 + 1) -CMA for Evolution Strategies," in *GECCO*, pp. 453–460, 2006.

[169] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing.," *Science (New York, N.Y.)*, vol. 220, pp. 671–80, May 1983.